SALIENCY AND EYE MOVEMENTS IN THE PERCEPTION OF NATURAL SCENES

Thomas Foulsham, BSc.

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

July 2008

Abstract

Humans inspect the environment around them by selecting a sequence of locations to fixate which will provide information about the scene. How are these locations chosen? The saliency map model suggests that points in the scene are represented topographically and that the likelihood of them being fixated depends on low-level feature contrast. This model makes specific predictions about the way people will move their eyes when looking at natural scenes, although there are few experimental tests of these predictions.

The experiments described in this thesis show effects of visual saliency on the likelihood and the speed at which objects are fixated. Experiment 1 shows that the potency of salient objects is moderated by the task being performed. When the task does not constrain the regions of interest, as in a general encoding situation, the saliency model performs better than chance estimates (Experiments 2 and 3). There are also sequential patterns of eye movements in this task—scanpaths—that the model does not reproduce. In visual search, participants can saccade to a target object, and this is quicker, in some cases, if the target is more salient (Experiments 4-7). A salient distractor impedes search more than a non-salient one (Experiments 8 and 9). The context of the scene also has an effect on search, and features of the layout, in particular the horizon, may cause an asymmetry in saccade direction (Experiment 10). Findings from research with a visual agnosia patient are consistent with the idea that scene understanding and saliency combine in guiding the eyes (Experiment 11).

These experiments support a framework that incorporates a task-driven prior, gist and the relevance of each region to the task, in addition to bottom-up saliency. Thus saliency is just one part of the way in which people move their eyes.

Publications

Parts of this thesis have previously been published as the following articles:

Chapter 3:

Foulsham, T. & Underwood, G. (2007). Can the purpose of inspection influence the potency of visual saliency in scene perception? *Perception*, 36, 1123-1138.

Chapter 5:

Foulsham, T. & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, 8, 2, 6, 1-17. http://journalofvision.org/8/2/6/

Chapter 10:

Foulsham, T., Kingstone, A. & Underwood, G. (In press). Turning the world around: patterns in saccade direction vary with picture orientation. *Vision Research*.

Acknowledgements

I am grateful to all of the people who have influenced the work in this thesis: my supervisor, Geoffrey Underwood; my colleagues in Nottingham (particularly the past and present occupants of Room 308); all those participants who have allowed me to screw an eye tracker onto their head; and CH. Some of the work in Chapters 11 and 12 was completed whilst visiting the University of British Columbia, which was made possible by the support of Alan Kingstone, Jason Barton and their colleagues. I am also grateful to those who have reviewed my work, and to the researchers at I-Lab at the University of Southern California for making the saliency map model freely available.

I would also like to thank those who have taught me where to look: my friends, family and particularly Anna, who have supported me in my studies and everything else.

Table of Contents

1 Introduction

1.1 Overview

This thesis is concerned with where people look when inspecting a scene. Since the first reports of eye movement recording in humans, researchers have been attempting to identify regularities in the way people move their eyes when viewing images. As I will show, people move their eyes systematically in order to select different parts of their environment and process these regions more efficiently. This research attempts to understand how these locations are selected and in doing so will provide information about how the brain integrates visual information and ongoing behavioural demands. A good model of eye movements will be able to predict where people will fixate in a scene, and this knowledge will reveal details about how vision works in the real world. When human resources are stretched, the places people look, or fail to look, have important consequences for what they remember and understand, and for the way in which they behave.

I will begin by introducing eye movements as an index of attention and an essential component of dynamic vision. Early research into scene perception revealed some interesting effects of the semantic components of the scene and of the task the viewer is performing. More recently models have been developed which predict eye movements based on the features of the image. This thesis concentrates on one such model—the saliency map model—and the second half of this introduction reviews this model in some detail. The subsequent chapters test the predictions of this model, with a view to improving the model and revealing patterns in the way people view naturalistic scenes.

1.2 Visual attention, eye movements and active vision

Vision is not a static process. In the normal environment, humans move their eyes several times each second, and these saccades are often described as ballistic, in the sense that they are fast and that their target is planned in advance. Between saccades, eye position is relatively stable. The resolution of the visual system is greatest at the centre of gaze—due to the density of receptors found at the fovea—and acuity decreases steadily as

the stimulus is moved further away from fixation. Moving the eyes allows the fovea to be directed at different parts of the environment, and knowledge of the regions and objects in a scene is built up over a series of these fixations. In addition to saccades the eyes may move to track a moving object (pursuit eye movements), to change the relative position of the two eyes (vergence eye movements), or in response to head or body movements. This thesis will explore only saccades, the mechanism by which people sample different areas of the visual field with foveal vision. Saccades are extremely fast, reaching a peak velocity of about $500^{\circ}s^{-1}$ (Becker, 1991). The duration of the relatively stable fixations in between is believed to depend on the type of stimulus and the reason it is being inspected: estimates of the mean fixation duration range from around 150 to 300 ms, though there is considerable variability between and within observers (Rayner, 1998). When possible humans and many other animals move their eyes in this way almost all the time. Even when eye movements are disabled, scanning head movements arise to select regions of the visual field in a similar way (Gilchrist, Brown, & Findlay, 1997).

Given the ubiquity of eye movements it is perhaps surprising that experimental investigations of attention have tended to focus on a different form of selection: covert attention. This is the term given to the preferential processing of a certain location or object while the eyes remain fixated and it is often visualised as a spotlight that selects certain parts of the visual field (Posner, 1980). This type of attentional selection is covert and unobservable, in contrast to eye movements, which can be labelled overt attention. The myriad findings of experimental research into covert attention will not be reviewed in detail here, but a few relevant results will be mentioned to illustrate some of the divisions identified within this concept. Firstly, experimenters have distinguished between attention that is exogenous, guided relatively automatically by stimuli outside of fixation, and that which is endogenous and deliberately shifted by the observer. Secondly, visual search experiments have probed the efficiency of observers searching for a target amongst distractors. This paradigm, and Triesman's feature integration theory in particular, has led to a distinction between processing which is achieved by serial shifts of attention which select items in a set one-by-one and the extraction of information from many items in the visual field in parallel (Treisman & Gelade, 1980). Several experiments in this

thesis are concerned with visual search and some of the theories put forward to explain these tasks are discussed in more detail elsewhere (see Chapter 8).

How are overt and covert attention linked? The two processes can be dissociated by paradigms that show changes in the locus of processing while fixation is stable or at a time scale too quick for a saccade to take place. On the other hand, it is increasingly acknowledged that covert shifts may be rare outside the laboratory, and that when they do occur they usually precede an eye movement. The enhanced processing of attended regions tends to also occur at locations to which an eye movement is planned and the two processes—attending covertly and planning an eye movement—may even be one and the same. This is the basis of the premotor theory of attention (Rizzolatti, Riggio, Dascola, & Umilta, 1987). The key paradigm of visual search has now been well studied with the recording of eye movements. The distinctive search slopes in this task, which are key for some theories of attention, can be related to the number of fixations made, assuming that several objects are processed in parallel during every fixation (Zelinsky & Sheinberg, 1997). Evidence such as this has led to the approach known as active vision, which argues that vision and visual attention is best understood by considering eye movements (Findlay & Gilchrist, 2003).

Attention is classically seen as a filter, a psychological bottleneck whereby the human information-processing system selects certain signals at the expense of others (Broadbent, 1954). In the real world, eye movements are a crucial processing limitation in that they are executed serially, but they allow processing to be concentrated on certain regions in a background with a huge amount of visual information. Thus, eye movements can be seen as a prototypical evolutionary development that maximises the useful signal in an environment that contains much more information than it would be computationally possible to process. While covert attention and its relationship with overt attention continue to be interesting, for application to natural behaviour models of attention must consider eye movements. Unless constrained otherwise, people tend to shift their eyes with their attention, and therefore this thesis will largely concentrate on eye movements as an index of visual attention. As Henderson (2003) argues, the measurement of eye movements provides a relatively unobtrusive measure of ongoing visual and cognitive

processing which is well suited to the studying of scene perception. Natural scenes are complex and the details that are important for the viewer often require the high resolution afforded by foveal inspection in order to be understood. The next section considers research into how people control their gaze within such scenes.

1.3 Early research into eye movements in scene perception

In his review of eye movements in information processing tasks, Rayner (1998) notes that many basic observations about eye movements were made before 1920. Saccades were observed and measured and some phenomena, such as the suppression of visual processing during saccades, were discovered during this time. Relatively few scientific studies of eye movements occurred from this time until the 1970's, due to the difficulty of accurately tracking the eyes. However, several classic studies in scene perception were carried out before 1970. In one of the first, Buswell (1935) showed subjects a range of artworks and observed some general patterns of eye movements. Some of these pictures were complex scenes and buildings, naturalistic stimuli far removed from the simple arrays generally used in the psychology and psychophysics of attention. Rather than being random, Buswell found that fixation locations were fairly consistent across participants and seemed to relate to the information in the picture. Fixations were concentrated on areas of detail and on objects and people rather than on the background.

Yarbus (1967) was one of the first to reveal the cognitive and task-specific nature of eye movements. He recorded eye movement patterns whilst people viewed scenes to answer various questions, and he showed that gaze patterns varied when participants were given different tasks to complete. Thus Yarbus emphasised that where people look will be different depending on the task they are performing. This simple fact will be of general importance for this thesis as it places limits on the degree to which the stimulus determines where people fixate. Since Yarbus, eye-tracking equipment has become much more sophisticated, allowing greater spatial and temporal resolution. Modern implementations are also much more portable and so the kinds of task studied have become more varied (Hayhoe & Ballard, 2005).

Published in the same year as Yarbus, Mackworth and Morandi (1967) compared the density of viewer fixations in each of 64 regions with the "informativeness" ratings of the regions made by a separate group of participants, in an attempt to quantify the information which people were selecting with their eyes. The number of fixations made in each region was positively related to its informativeness. What constitutes "informativeness"?

One way of considering this is to look at the effects of violations in the context of a scene. Loftus and Mackworth (1978) presented their subjects with line drawings of natural scenes and tried to manipulate the informativeness of an object in it. They reasoned that an object that is semantically inconsistent with the rest of the scene (one which contravenes the "gist") is more informative and so should attract more fixations than one that is expected. The researchers embedded line drawings of various objects in different scenes. In half of these scenes, the object was consistent (a tractor in a farmyard) while in the other half it was inconsistent (an octopus in the same scene). Sure enough, participants who explored the scene were more likely to move their eyes to an inconsistent, informative object than to an object that was consistent. Significantly, this difference was found after just one fixation on the scene, and target saccades were large in amplitude, suggesting that anomalous objects could be identified as such with only peripheral vision. The context of the scene was available quickly enough to influence eye-guidance and prompt large saccades straight to the informative area. The presence of scene gist and layout appears to be available in the time frame of the first eye movement: it can be reported after very brief exposures (Potter, 1976) and seems to boost object recognition (Biederman, Rabinowitz, Glass, & Stacy, 1974) so might affect eye movements in this way.

Unfortunately, subsequent investigations failed to reproduce the results of Loftus and Mackworth (1978). De Graef, Christiaens and d'Ydewalle (1990) found that implausible objects (such as a dumper truck on a roof) were not fixated earlier than other objects, and Friedman (1979) failed even to find a difference in fixation density between congruous and incongruous regions. Henderson, Weeks and Hollingworth (1999) used two tasks (one where observers tried to memorize the scene, and one where they were

searching for a target object) while investigating eye movements over line drawings containing objects which were either consistent or inconsistent. In the first task, Henderson et al. (1999) failed to replicate Loftus and Mackworth's key result: inconsistent objects were not fixated any earlier or from further away than consistent objects. In the search task, where participants were given an object label and asked to indicate whether or not it was present in the scene, consistent objects were detected faster and fixated sooner than inconsistent objects. Henderson et al. explained this result, as being caused by expectancies of where consistent objects might be found: searching for a semantically consistent object is easier because it has a probable location within the scene.

One reason for the failure to replicate this result might be that the anomalous objects used by Loftus and Mackworth (1978) also differed in terms of how much their visual features stood out. Underwood and Foulsham (2006) used similar tasks to those of Henderson et al. and showed that visual distinctiveness and incongruity interact: out-of-place objects only preferentially attracted attention when the object was inconspicuous. Based on their findings, Henderson et al. suggested that initial fixation placement is based on visual factors, rather than being informed by a semantic interpretation of a scene. Their framework suggested that an early parse of the scene identifies "blobs" of potential interest, which are then evaluated in a "saliency map". This map represents the positions and relative informativeness of scene regions, and saccades are directed to the most salient regions. Crucially, saliency is determined by purely visual properties such as brightness, and it is only after regions have been fixated that their saliency can be altered due to semantic inconsistency or other cognitive factors, allowing them to be fixated for longer, or refixated if of particular interest. The remainder of this introduction will consider explicit models of saliency that aim to predict where attention should move in natural scenes.

1.4 Visual saliency in scene perception

This thesis is about the effect of visual saliency on eye movements in scenes. Following the framework of Henderson et al. (1999) I will use the term *visual saliency*, or just *saliency*, to refer to a property of a point in a scene, which makes it likely to be fixated

and which is independent of the observer's knowledge or task. In other words, the experiments are concerned with the stimulus-driven or bottom-up factors involved in eye movements in natural scene perception. Although saliency (often used interchangeably with *salience*) is sometimes used in a more general sense to describe anything that is pertinent to a task or which attracts attention and fixations, I will use it in only the sense of a specific bottom-up determinant of attention related to the particular model described below. As I shall demonstrate, among the visual properties believed to be important are intensity differences (how bright something is relative to its background) and contour information signifying the location of objects or textures. These variables are assumed to be low-level: they do not require knowledge of what the scene is or recognition of the to-be-fixated object; they are computed by brain areas which are early in the visual pathway; and they may guide the eyes in an automatic or exogenous way.

As Buswell (1935) noted, some areas are more likely to receive inspection than others and at its simplest the bottom-up approach involves determining what these areas have in common, without recourse to the cognitions of the observer. The problem at hand is how the eyes are directed to regions of interest given the limited resolution in the periphery, and it is plausible that low-level features might resolve this task, leaving processing resources free to deal with recognising and understanding what is at fixation.

1.4.1 Comparing scene statistics with fixation locations

Mannan and colleagues were among the first to compare the locations of specific visual features within a scene and those of the fixations made when this scene is viewed. An initial report by Mannan, Ruddock and Wooding (1995) suggested that the places where people looked in monochrome natural scenes did not vary very much when these scenes were high- or low-pass filtered. This filtering made identifying the scenes very difficult, so it was argued that the distribution of fixations must be determined by local image features which were relatively unaffected by the global image modifications. The authors subsequently compared the fixated locations with: the locations where luminance was highest; those where luminance was lowest; those where Michelson contrast was highest; those where edge density was highest; and those with the greatest density of high spatial

frequency content (Mannan, Ruddock, & Wooding, 1996; see also Chapter 4 for some details of the comparison metric used). Of these features, only locations of image contrast and edge density were reliably similar to the fixated locations.

By comparing measures taken at regions that have been fixated with those taken at unfixated, or randomly generated regions, several other researchers have recently argued that low-level features guide eye movements. Reinagel and Zador (1999) showed that fixated patches had a higher luminance contrast than randomly selected regions, as did Parkhurst and Neibur (2003). Tatler, Baddeley, and Gilchrist (2005) used a signal detection method to assess which features are indicative of fixated regions. The researchers demonstrated that complexity at high spatial frequencies and the presence of edge and contrast information are distinctive of regions that are selected by overt attention. Interestingly, this may depend on the length of the preceding saccade (Tatler, Baddeley, & Vincent, 2006). Other researchers have extended the scene statistics approach to look at second- and higher-order image statistics (Frey, Konig, & Einhauser, 2007; Zetzsche, 2005) These features are indicative of corners, junctions and curved lines or edges, and nonlinear filters are necessary to detect them.

A problem with the scene statistics approach is that it identifies correlates, rather than causes, of attentional selection. If the eyes are guided by a top-down signal (for example a desire to look at objects or to move the eyes to a particular head-centred location) and if certain visual features tend to be higher at these locations then a correlation will occur without telling us anything useful about what caused the eye movement. This is a major impetus for doing controlled experiments and thus for the work in this thesis. A particular issue with the correlational nature of some of this research concerns which areas the fixated regions are compared to. If they are compared to random locations in an image, then the resulting differences may just be showing that fixations and the salient feature in question have a similar distribution, but not that one causes the other. It is necessary, therefore, to correct for the distribution of fixations, and particularly for the tendency shown by many observers to fixate in the centre of an image. Mannan et al. (1995) and Tatler et al. (2005) address these issues to some extent and I consider them in more detail in Chapter 5. A useful way to combine the different image

features that may contribute towards attentional selection is to produce an explicit model of saliency that produces predictions. Testing these predictions can avoid some of the problems of a correlational, image-based approach, and this is one of the aims of this thesis. The next section introduces just such a saliency model.

1.4.2 The saliency map model

Koch and Ullman (1985) suggested that attentional selection was accomplished using a "saliency map" which explicitly encodes the saliency of each part of the visual field. Attention selects the point with the highest saliency (a "winner-take-all", WTA algorithm), which is then suppressed ("inhibition of return", IOR) so that attention can be disengaged and moved to the next most salient point. The concept of a distributed, topographically organised map is common in models of both covert attention and eye movements (e.g. Findlay & Walker, 1999; Wolfe, 1994). This representation is a simple way to explain how attention moves over a scene, although it has proved difficult to model exactly how information from different features is combined into a single map.

Itti and Koch's (2000) bottom-up model of visual attention is a computational implementation of a saliency map mechanism that guides attention to select salient regions of an image or scene. As such it is the most well specified model of stimulus driven attention in natural scenes and will be discussed in detail. Figure 1.1 shows the different stages of the model, as applied to an example scene. The specific implementation of the model used to derive this figure, and throughout this thesis, is outlined in the general methods in Chapter 2. For a more detailed description of the algorithms at each stage, the reader is referred to Itti, Koch and Neibur (1998) and Itti and Koch (2000).

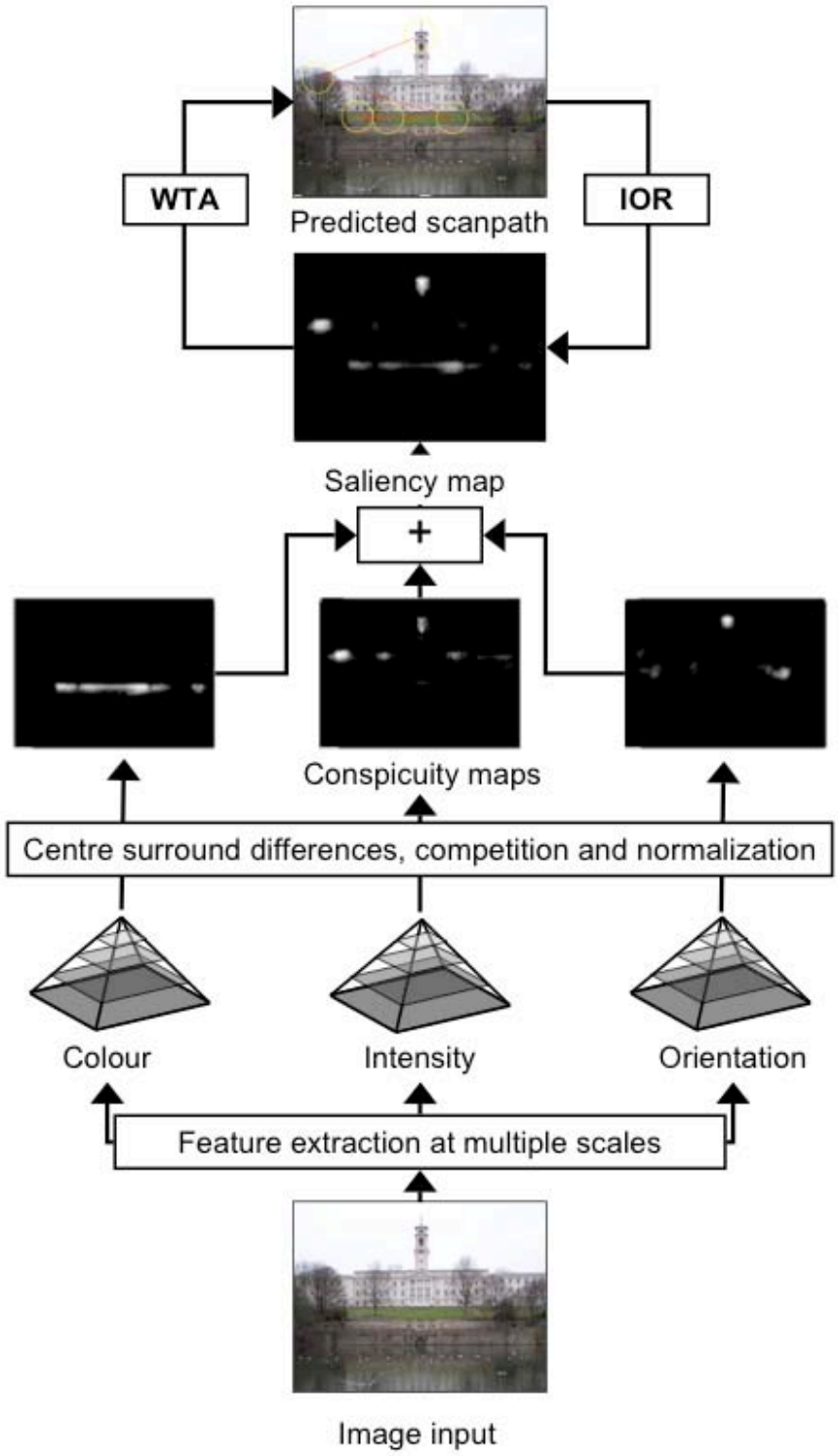


Figure 1.1. An overview of the saliency map model described by Itti and Koch (2000), as applied to a natural image. The image provides the input at the bottom of the model. Linear filtering extracts variations in colour, intensity and orientation across several scales using Gaussian pyramids. These are combined to give centre-surround contrast within each feature, and the features are then summed into a single saliency map.

The model is based on three feature dimensions: colour, orientation and intensity. These features are extracted from a colour image that represents the input to the retina, using a range of linear filters that act in parallel to simulate preattentive processing. This filtering occurs at several different spatial scales, created by Gaussian pyramids that sample the image at a progressively coarser scale. In Itti and Koch (2000) 8 different spatial scales are used, with image reduction factors ranging from 1:1 to 1:256. In order to produce a centre-surround structure, the model computes the difference for a given feature between high and low spatial frequency levels of the pyramid. This results in a system that responds most to local spatial contrast and so replicates receptive fields in the early visual system. One feature type encodes the intensity contrast of the image by comparing centre intensity and surround intensity. Colour information is first normalised by the intensity channel and then processed using a red/green channel and a blue/yellow channel. In each case, a double-opponency calculation is carried out between the centre and the surround, with the absolute value of the difference as the resulting output. Four channels encode orientations of 0, 45, 90 and 135˚. Oriented Gabor filters extract orientation values and the centre-surround differences are computed to give a measure of average local orientation contrast.

The result of the preattentive feature extraction stage is a group of feature maps within each of the seven channels (intensity, red/green, blue/yellow and four orientation channels). In each case, the maps represent the local differences within each feature, giving a measure of the discontinuity at any one location. Combining a large number of these maps into an overall saliency map is the challenge faced by the remainder of the model. This challenge consists of specifying how to compare or weight different feature values and how to maximise the signal to noise ratio in order that a region which is highlighted in one map is not lost amidst similar activation at other locations in many other maps. Itti and Koch propose that the best way to overcome these problems is to simulate the long-range interactions thought to be present in the brain. Feature maps are rescaled to a fixed range to eliminate feature specific value differences. They are then iteratively convolved with a two-dimensional Difference-of-Gaussians (DoG) kernel that acts to create spatial competition for salience within each map. The DoG filter gives

strong excitation locally but broader inhibition from regions further away. After several iterations, maps with many highly activated regions are suppressed (e.g. one with input from a stimulus with many similar elements), while maps with only a few significant peaks are accentuated so that original maxima are potentiated (e.g. one with a stimulus containing one distinctive element). Following further normalisation, the maps from each spatial scale are summed to give a single "conspicuity map" for each feature type. These maps can be thought of as highlighting the location of the most significant discontinuities in the stimulus. For example, the conspicuity map for the colour channel shows where the biggest local changes in colour can be found, and so identifies regions where a patch of colour contrasts with the background and stands out. As an additional step, the three conspicuity maps undergo further iterations to enhance within-feature competition, before being summed into the unique saliency map.

The saliency map is a scaled topographic representation of the image whose maximum represents the most salient location according to the computation outlined above and which attention should select. In order to identify this location so that it may be selected by covert attention or become the target for a planned saccade, a WTA network is applied. This network is implemented as a layer of integrate-and-fire "neurons" with strong global inhibition. With input from the saliency map, the most salient location will be the first neuron to fire and causes an attention shift and inhibition of all the cells in the layer. The final process necessary for the model is a simulation of IOR to prevent the same most salient location from always being the "winner" in a static image. IOR in covert attention has been demonstrated in many experiments, as well as sometimes with eye movements (see Klein, 2000, for a review). The model implements IOR as inhibitory feedback from the WTA network to the saliency map that takes the form of a DoG kernel, transiently inhibiting the winner and its neighbours. As a result of the shape of this filter, conspicuous regions closer to the original winner will become slightly more salient, an implementation designed to promote local shifts of attention, rather than long-range ones.

The computational detail inherent in this model has several advantages for investigating scene perception. Firstly, it is sophisticated enough to use complex

photographs and does not rely on the simplified line drawings or stimulus arrays common in attention research. Indeed it could be argued that some of the discrepancies in the results discussed above may arise from failing to take into account stimulus differences. While model building should and must begin by looking at limited, simplified examples of real world tasks, sooner or later these models must be anchored in more complex implementations so as to allow them to generate useful predictions. A second advantage of the model, then, is that it produces specific predictions that can be compared to the eye movements of human observers. The aim of the model is to predict attentional selection: which regions will be fixated due to bottom-up factors and in which order? The next section will consider a third advantage of Itti and Koch's model, that it is biologically plausible, before assessing attempts to evaluate it with natural scenes.

1.4.3 Neurobiological evidence for a saliency map

Itti and Koch (2000; see also Itti & Koch, 2001) justify their model with reference to what is known about the neural systems involved in visual attention. Preattentive coding for discontinuities begins in the retinal ganglion cells, where receptive fields are organised in a centre-surround fashion. This makes them particularly responsive to edges in the luminance profile. Neurons in the visual cortex continue to show this structure, emphasising the differences between centre and surround regions of the receptive field in various stimulus dimensions including orientation and direction of motion. This processing is reflected in the model by the centre-surround computations used to combine spatial scales into a single feature map. The coding for colour in the first stages of the visual pathway inspires the double-opponency found in the model. Other aspects of the model directly imitate the kinds of filtering that goes on in vision. Convolution with a DoG style filter and Gabor functions, which resemble biological impulse response functions, are two examples. The early feature extraction in the visual system happens in a massively parallel fashion across the visual field and is relatively unaffected by attention. Preattentive feature extraction seems to rely on contextual modulation not just locally within the classical receptive field but over long-range connections in V1. For example, the responses of orientation-selective cells are enhanced when they extend into contours outside the receptive field (Gilbert, Ito, Kapadia, & Westheimer, 2000). Such

long-range interactions are the inspiration for the within feature competition implemented in the model.

There are significant difficulties with recording neural responses to saliency as opposed to simple feature properties or presence. However, several studies have identified neurons that are selective for visual saliency in areas such as the frontal eye fields, the pulvinar and multiple parts of the parietal cortex of monkeys (see Treue, 2003 for a review). For example, Gottlieb and colleagues (Desimone & Duncan, 1995; Gottlieb, 2002) found neurons that responded vigorously to a stimulus when it was salient, but not otherwise. When the stimulus flashed rapidly outside the receptive field before being saccaded to it caused more activity than one that appeared straight into the receptive field or one that had not been flashing. The fact that multiple locations for a saliency map have been reported suggests that one unique map may not be present. Some authors have argued that saliency is not coded explicitly but is computed implicitly by distributed modulatory activity in feature maps (Desimone & Duncan, 1995) while Li (2002) proposed that the saliency map can be generated directly from firing rates in V1. Although originally formulated as a theoretical construct, the concept of a saliency map may be closer to the neural reality than first thought.

1.4.4 Tests of the model

To evaluate a saliency map model it is necessary to compare the regions selected by the model as salient (predicted fixations) with the regions that are in fact fixated. Itti and Koch (2000) set their model several tasks. Firstly, the model replicated classic visual search behaviour in that it took longer and made more attention shifts in conjunctive versus feature search. Thus a red horizontal bar amongst green horizontal bars was selected very quickly (as in "pop-out" effects in humans) while a red horizontal bar amongst red vertical bars and green horizontal bars took longer to be selected and showed a linear trend with an increasing numbers of distractors. Although this thesis focuses on scene perception rather than simple visual search, Chapters 8 and 9 investigate the effects of saliency in a visual search for objects.

The visual search simulation was important due to the large amount of normative data in humans available, but it is harder to find an objective performance criterion for

natural scenes. Itti and Koch compared model performance to human search times when looking for small military vehicles in very high resolution outdoor scenes (Toet, Bijl, & Valeton, 2001). No additional "training" or feature weightings were used for the model simulations. The model consistently out-performed human observers and its focus of attention moved to the target region quicker than human searchers (who indicated that they had found the target with a button press). This analysis was conducted with the reasonable (though conservative) assumption that attention shifts take place at an average rate of three per second. While the superior performance of the model is interesting, given that the comparison involved only a relatively crude measure of response time and not actual fixation data the conclusions that can be drawn are limited.

It is worth summarising the results from Itti and Koch (2000). Although the model was completely bottom-up (it had no higher-order knowledge of what it was looking for) it could locate targets as well or better than humans in both a simple array of oriented bars and a very complex natural search task. Although proposed as a model of search, it was taken as advantageous that there was no task specification so that the model could predict "free" viewing as well. The authors were cautious about how well the saliency map model would predict actual eye movements, preferring to focus on covert attention and pointing out that the model contained no simulation of eye movement mechanics; it did not address the literature on attention capture by motion or abrupt onsets (Hillstrom & Yantis, 1994); it did not deal with visual search phenomena such as search asymmetries or spatial grouping effects (Treisman & Gormican, 1988); and most importantly, it was oblivious to any scene or task context.

Nevertheless, further research has attempted to quantify the comparison between a saliency map and fixation locations. Parkhurst, Law and Niebur (2002) compared the strength of activity in a saliency map (the model-predicted saliency) at the fixated locations of four subjects with that expected by a uniform fixation distribution. The resulting "chance-adjusted salience" was significantly positive, indicating a greater than chance probability of fixating regions with high salience. There were two other noteworthy findings. The chance-adjusted saliency was higher for artificial fractal images than for natural scenes, implying a greater correlation with fixations. In addition,

when looked at over time the relationship seemed to decrease such that saliency was highest at the locations of the first few fixations and declined as viewing went on.

Additional studies have altered various aspects of a basic saliency map model in order to assess the contribution of different features to eye guidance. For example, Hugli, Jost, and Ouerhani (2005) suggest that the influence of colour (as opposed to just luminance) is significant, as is the influence of depth information. Itti and colleagues have reported that an added motion channel contributes more to the successful predictions of a saliency map than colour and other features do in dynamic scenes (Itti, 2005).

In a very thorough paper and perhaps the most complete evaluation of the saliency model to date, Peters, Iyer, Itti and Koch (2005) used a similar method to Parkhurst et al. (2002) but with a variety of slightly different models. The human data came from 12 participants whose eye movements were recorded whilst they performed one of two tasks: "free viewing" (in fact, they had to indicate the half of the image they found more interesting) and contour-detection (where the task was to respond as to whether a highlighted contour was present in the previous image). Several classes of images were used: overhead satellite images; greyscale photographs of outdoor scenes; colour fractals; and greyscale Gabor arrays. To analyse the relationship between the observed scanning patterns and the activation in the saliency map, the authors normalised the maps to give them a mean of zero and a standard deviation of one, as in a z-statistic; calculated the mean value at fixated locations; and compared this mean to the distribution of normalised values in the image. The "baseline" saliency model resulted in a normalised scanpath salience that was significantly greater than zero for all the complex images (the Gabor patterns were used only as a control and led to chance performance). Thus, as in Parkhurst et al. the relationship between the bottom-up model and the fixation locations was closer than expected by chance. Adding various components such as extra short- and long-range interactions within the orientation channel improved the fit slightly. A convenient way to express the success of the saliency map model in this experiment is as a percentage of the inter-observer relationship. This can be represented as a fixation density map, and the extent to which an observer's fixations can be predicted from such a map formed from the remaining N-1 participants provides an upper limit on what a purely

bottom-up model can predict. Peters et al. report that the saliency map model explains around 50% of this variation.

I have previously mentioned, and will return to, the problems of comparing saliency at fixation with its uniform distribution within an image or class of images. Peters et al. (2005) improve on the method of Parkhurst et al. (2002) by tailoring their calculations to the distribution of each image and thus taking into account the variation in saliency present in different scenes (by normalising the saliency distributions). However, the inference in this research is still correlational rather than causal. Given that one of the tasks in Peters et al. was to look for "interesting" regions the correspondence between human- and model-selected locations might be due to the observer actively seeking out these regions rather than them driving attention. The particular issue regarding the centralisation of fixations and saliency is addressed to some extent by Parkhurst et al. and Peters et al., who introduce eccentricity-dependant filtering into their models. They argue that this bias can, to some extent, be modelled by weighting information closer to fixation more strongly than that further way, in line with the decline in sensitivity found in the visual system. Peters et al. show a large improvement in model performance with this addition. However, if there are constraints on the distribution of fixations that do not arise from the distribution of image features (such as a motor bias towards the centre, or a cognitive expectation that interesting things are located near the middle of a scene) then these could explain some or all of the correlation between model and human fixations. Vincent, Troscianko and Gilchrist (2007) have recently shown that variable sampling of the image with greater sampling in the centre, as found in the retina, causes saliency map models to be less reliable.

Although Parkhurst et al., (2002) attempt to reduce the influence of task by using "free viewing" instructions, Hayhoe and Ballard (2005) have argued that the task being preformed is crucially important and that embodied models which take into account the goals of the "agent" are more useful than the notion of saliency. In support of this idea, several authors have argued that saliency can explain little or none of the variance in fixation locations in natural tasks such as walking (Jovancevic, Sullivan, & Hayhoe, 2006; Turano, Geruschat, & Baker, 2003). For example, in Turano et al. participants

walked down a corridor and were instructed to enter the fifth door on the right. The fixations made were better explained by coarse location information (e.g. a bias towards the right of the corridor) and features (e.g. vertical edges indicative of doorways) than by the saliency map model. The saliency map model has also been used in experiments looking at change detection, and changes to salient objects may be easier to detect in some cases (Stirk & Underwood, 2007; Wright, 2005).

At present there are very few experimental evaluations of a saliency map model, and this thesis aims to provide some. The next section briefly outlines some other (non-saliency based) models of eye movement control that are relevant to scene perception, and I will then discuss how recent updates attempt to combine bottom-up saliency with cognitive, knowledge-based factors.

1.5 Alternative models of eye guidance

Research which aims to formulate detailed models of eye movement control is dominated by those based on reading. A large amount of progress has been made because reading represents a rather closed domain in which there is a large corpus of data showing trends in where (within words) people fixate and for how long, given the lexical and semantic properties of the words and the sentence (see Rayner, 1998, for review).

There are far fewer detailed models of eye movements in visual search and scene perception: the stimulus in these cases is more complex to display; has many more dimensions that can vary; and permits movements in two dimensions (whereas the saccades in reading are basically one-dimensional). In terms of controlling when to move the eyes, a model similar to the reading model proposed by Morrison (1984) has been suggested. In such a model, the trigger to move the eyes comes once the information at fixation (e.g. an object) is identified. Some of the effects observed in reading have been replicated with objects or scenes (e.g. the preferred viewing position, Henderson, 1993; see Chapter 9). Some of the tools from reading (e.g. gaze-contingent or "moving window" paradigms) have also provided information about what controls fixation durations in scene perception (see Chapter 7).

Findlay and Walker (1999) proposed a well-specified model of eye movements that aims to explain both when the eyes move and where they go. The model is based on

some of the pathways known to exist in the brain and it comprises a fixate centre and a move centre which are linked by lateral inhibition. This provides a plausible framework for processing demands to promote fixation and inhibit making a saccade. Activation in the move centre, conversely, results in disengagement from fixation and the programming and execution of a new saccade. The representation of where to move to is a topographical map which is similar to the saliency map of Itti and Koch (2000), although Findlay and Walker also use the term "intrinsic salience" for locations which are a priori more likely to attract saccades. Although I will mention fixation duration in some of the following experiments, the majority of this thesis is concerned with where people fixate, and as such I will now outline some models that concentrate only on this.

Several researchers have described models of eye movements that can apply to tasks involving natural scenes. Rao, Zelinsky, Hayhoe and Ballard (2002) built on the ideas from Wolfe's (1994) Guided search model to model the process of searching for natural objects in scenes. In this model, fixation locations are selected based on how well the features at each location correlate with a stored iconic representation of the target. This iconic representation is a vector of filter responses at different spatial scales. As in the saliency map model, all possible locations are represented in a map and the gaze moves to the point where the target is most likely to occur. Thus, this model is completely reliant on knowledge of the targets features, and does not depend on the a priori saliency of scene regions: the fixations are determined by the search task. With some further implementation details (such as the assumptions that an eye movement is made before the map is completely computed and that coarse features are processed earlier than finer ones) Rao et al. show that the model simulates eye movements rather well, accounting for some of the phenomena commonly observed such as "centre-of-gravity" saccades where the eyes land between two points of interest.

An alternative to the notion of saliency in saccade target selection is the information-theoretic approach described by Renninger, Verghese and Coughlan (2007) and Raj, Geisler, Frazor and Bovik (2005). In this approach, the visual system selects the locations which give the most information for a task, or which reduce the amount of "uncertainty". Renninger et al. investigated a learning task for simple contoured shapes,

and they modelled the information available at each point in terms of the complexity of the contours present there (the orientation information in a region with straight lines is relatively certain, whereas bumpy contours contain more information). Their model, which assumes that each fixation is planned to maximise the available information, was a closer fit to human data than the Itti and Koch (2000) model (although the latter still performed slightly above chance). In natural scenes, Raj et al. devised a model that generates a scanpath based on local contrast information and an algorithm that minimises the total entropy, or uncertainty. Although this model was not directly compared to real eye movements this is a nice way to generate a sequence of fixations that accumulate information about an image, and it has the advantage of requiring few additional parameters and no specific task or target.

1.5.1 Bottom-up, top-down or a mixture of the two?

The approaches discussed thus far can be roughly grouped as bottom-up or top-down models. Bottom-up factors depend on the stimulus and are therefore constant across observers and tasks, and saliency is one example. Of course, the information in the stimulus will be required for most tasks, so that objects can be identified for instance, but it may not be the main determinant of where the eyes fixate.

The alternative is guidance by top-down factors. I will use this term for models that depend not just on local stimulus attributes but on the task and cognitions of the observer. As a result, these models must be invoked to explain the difference in eye movements associated with different tasks (Yarbus, 1967). For example, of the models above, Rao et al. (2002) is a top-down model because it relies on the observer's task (to search) and on their representation of the target. Changes in the target would lead to different fixations despite the external information remaining constant. A more controversial example of a top-down model is Noton and Stark's (1977) scanpath theory, which is discussed in detail in Chapter 4. While the bottom-up versus top-down approach may be useful as a rough framework, it is unclear whether "top-down" should be defined phenomenologically (as a subjective sense of control), psychologically (as a state induced by task instructions) or physiologically (as feedback from "higher" to "lower" brain

areas). These definitions are not necessarily compatible (Frith, 2005) and so I will remain cautious with this terminology.

The research discussed thus far demonstrates that saliency map models do at the very least describe image properties that are important in determining gaze. Of course, when changes in task lead to different scanning patterns over the same stimulus, saliency models are fundamentally unable to provide a full explanation and so some models have been proposed which attempt to combine both bottom-up and top-down factors in eye guidance.

Navalpakkam and Itti (2005) updated the saliency map model to incorporate top-down knowledge of a target. The authors aimed to weight the purely bottom-up saliency map with information about the features present in the target (as in Wolfe, 1994 and Rao et al. 2002). This is achieved by weighting the different feature channels according to relevant features from the target (where a relevant feature is one that has a high mean activation and a low variance). For example, searching for a horizontal line at a certain scale will increase the weight of the 0˚ channel as well as its parent, the orientation channel. The weighting of irrelevant channels (such as colour in the previous example) will be reduced. Thus the saliency map is biased to form a "task-relevance" map that then controls fixations as in the pure saliency model. The authors show that the model can model search efficiently, in principle, and that it does better than the template-based model of Rao et al. in the case of feature versus conjunction search. Given a target template, a purely top-down model of the sort proposed by Rao et al. would predict all targets would pop out, when in fact conjunction search targets do not. Even without a known target Navalpakkam and Itti's model can predict pop-out (in this case the weightings are equal for all channels and the model performs just as the pure saliency map model). Thus this model is perhaps more flexible and can be applied to additional tasks beyond just searching.

Zelinsky and colleagues have combined the Rao et al. (2002) top-down model with one that is purely bottom-up, using various different mixtures of the two (Zelinsky, Zhang, Yu, Chen, & Samaras, 2005). Basically, a bottom-up saliency map is computed (as in Itti and Koch, 2000) and this is combined linearly with the top-down map showing

the correlation of the local image features with the vector of features present in the target. Zelinsky et al. tested 5 different combinations of the two maps and the results were dramatically in favour of top-down guidance: the best performance was a model which had no contribution from the raw saliency map, and even a 25% contribution of bottom-up information led to significantly worse performance in terms of the number of fixations to find the target. Chen and Zelinsky (2006) recently described a simple object search task, showing that top-down knowledge dominates (see also Chapter 9).

In these models, target knowledge is used to weight the features contributing to the saliency map in favour of those possessed by the target. An alternative way that the top-down cognitions of the observer could have an effect is by enhancing regions of the map corresponding to the task-relevant locations (in cooperation with scene layout or gist). These two possible influences of a search task are similar to those included in Findlay and Walker's (1999) model of saccade generation as processes of spatial and search selection, where regions or features are boosted based on the target. Henderson, Brockmole, Castelhano and Mack (2007) argued that saliency cannot account for eye movements in a real-world search task involving counting the people in a scene. In addition, this study compared the scene statistics of fixated and non-fixated regions and found a difference (as expected from the research discussed previously). However, there was also a difference in how meaningful the regions were rated, and so the authors argued that saliency may not have played any role in determining where people look. In other words, the places where people look may be better predicted by their expectations about where things occur (e.g. that humans tend to be near the ground) than by saliency.

The idea of gist and spatial layout combining with our expectations about where targets will be has recently been combined into a model of contextual guidance by Torralba, Oliva, Castelhano and Henderson (2006). Torralba (2003) showed that gist can be acquired in a straightforward, holistic fashion from low spatial frequency features. The contextual guidance model effectively filters a bottom-up saliency map so that attention is directed only to the locations in a scene that are likely to contain the target, based on this gist representation and prior experience with object locations.

1.5.2 Rationale for choosing a model

I have now discussed several models designed to account for where people move their eyes in natural scenes. This thesis focuses on the purely bottom-up Itti and Koch (2000) model of saliency. To summarize its advantages: it is computationally precise and specifies all individual procedures; it produces specific predictions; it is based on features which are to some extent neurally plausible; and it has received support from studies of scene statistics and from direct tests of the model (e.g. Parkhurst et al., 2002).

In comparison to other possible bottom-up models, the one chosen makes relatively few additional assumptions. For example, it does not place any particular weight on any one feature in eye guidance. Other models are either not implemented computationally (e.g. Findlay & Walker, 1999), making them hard to specify with complex natural stimuli, or they require additional assumptions or learning based on a target (Rao et al., 2002, Torralba et al., 2007). The code for the saliency map model is freely available which means that it has been used by many different researchers, to which this thesis can be compared. As the most popular bottom-up approach the saliency map model is likely to produce very similar predictions to other saliency models, although one should be cautious transferring the present results to other models. Perhaps the biggest advantage of the model is that it can make predictions across different tasks, unlike the top-down models that are difficult to extend beyond the situation for which they were designed (normally visual search).

Since its publication (and since the work in this thesis was begun) the saliency map model has been regularly updated, and some of these additions have already been discussed (see, for example, Navalpakkam & Itti, 2005; Peters et al., 2005; Itti, 2006). In many cases the additional features change the model's performance only slightly, but I will speculate about how they might change my results where relevant. The specific implementation used is described in Chapter 2 and is identical to the standard model used in Itti and Koch (2000). This remains a thorough and testable model of the principles of eye guidance based on visual saliency.

1.6 Outline of this thesis

This chapter has reviewed some of the low- and high-level influences on eye movements in natural scene perception. Early research indicated that fixation positions were not random and were tied to informative content and task goals. Given the limitations of time (people move their eyes very quickly) and peripheral resolution, it is plausible to think that bottom-up features might be used to determine where to look. The success of computationally precise saliency map models gives an opportunity to investigate to what degree this is the case. However, there is very little experimental evidence for the predictive power of these models in natural scenes. This thesis describes experiments that look for a causal link between saliency, eye movements and behaviour, and this is largely accomplished by deliberately manipulating the saliency of objects or regions of interest. The strongest models argue that saliency can explain all fixations regardless of task and top-down processing. At the other extreme it has been suggested that some tasks, particularly visual search, are not affected by saliency.

Following a general description of the methods used throughout the experiments, Chapters 3 to 11 describe the findings of a number of experiments into the effects of saliency in scene perception. In Chapter 3, two tasks from the scene perception literature (memorising and searching) are employed to see how any influence of saliency varies with task. Chapters 4 and 5 expand the analysis in a scene recognition task to look at the similarity of scanpaths made in different conditions. Chapter 4 concentrates on different algorithms for comparing sequences of fixations, and Chapter 5 uses these methods and a range of other analyses to evaluate the spatial and sequential predictions of the saliency map model. This chapter also explores some interesting findings in terms of the degree to which scanpaths are recapitulated on multiple viewings, and it highlights the importance of different systematic biases in where people look across a range of scenes. Chapter 6 uses the model to screen image regions that are then probed in both a memory and a search task. Observer performance for salient and control regions is compared, and this procedure is extended in Chapter 7 where several gaze-contingent conditions manipulate the visual information present in the scene. Chapters 8 and 9 return to some of the questions raised in visual search by looking at the effect of target and distractor saliency

in simple arrays of realistic objects. One of the recurring themes is the extent to which saliency effects cause or are confounded by some of the biases seen in eye movements. Chapter 10 looks at one of these biases—the tendency to make horizontal saccades—in some detail. Finally, Chapter 11 reports what was found when some of the experiments were performed with a patient who had visual agnosia.

This thesis aims to explore some of the predictions of the saliency map model in scene perception, and in doing so reveal how the visual-cognitive system selects information to attend to. Discussing the findings in Chapter 12, I will argue that saliency has widespread effects in most tasks, but that these effects must be moderated by the context provided by the scene and the observer's cognitions.

2 General methods

This chapter outlines the methodology used in the experimental chapters that follow. In investigating the effect of visual saliency in natural scenes, Experiments 1 to 10 used the same implementation of a saliency map model and similar stimuli, apparatus and data analysis. For the sake of brevity these common methods will be described here, although where methods differ they will be included in the relevant chapter.

2.1 Participants

In order to test the saliency map model of eye movements, each experiment tested a sample of individuals from the general population. As far as possible the participants used in all experiments did not differ in any systematic way (with the exception of the experiments in Chapter 11, which were performed on a patient and several age-matched controls). All participants were drawn from a university student population, meaning that they were aged between 18 and 30, with a similar socio-economic and educational background. Males and females contributed to the samples in approximately equal proportions, and this thesis will not consider any effects of gender.

All participants reported having normal or corrected-to-normal vision. The eye trackers used were very tolerant of observers who wore eyeglasses, but where these led to an unacceptably high rate of errors or missing data, that participant was replaced. In all cases, participants were paid an inconvenience allowance to take part, which was equivalent to approximately £6 per hour. Before each experiment, participants gave their consent to take part, with the knowledge that they were permitted to leave at any time. On completing the procedure participants were debriefed about the project where necessary.

2.2 Stimuli

As far as possible, the experiments described in this thesis use complex, realistic photographs of scenes and objects in order to explore the guidance of attention and eye movements. While much research into attention has used relatively simple visual stimuli (such as T or L figures or lines of varying orientation) the saliency-based approach to eye guidance aims to also predict realistic eye movements. This is useful for applied contexts

such as research aiming to design machines that can identify objects in the real world. There is also evidence that visual processing is different for natural scenes than for more artificial stimuli (Braun, 2003). A final justification for using natural images is that it is often difficult to probe the top-down knowledge that humans possess about the world around them using materials that lack this context.

The exact criteria for choosing stimuli are described in each experiment. The stimuli were digital photographs of natural scenes and objects that might be encountered in the real world, captured using a 5 megapixel digital camera or obtained from commercially available collections with a similar resolution. In Experiments 6, 7 and 9 photographed objects were separated from their backgrounds and arranged in arrays using the software Adobe Photoshop. This program was also used to filter and rotate the pictures in Experiment 5. The stimuli were not manipulated in any other way and were all presented at a resolution of 1024 by 768 pixels. In the following experiments stimulus measurements will be expressed in degrees of visual angle.

It has been reported that faces and other biological stimuli attract attention in a relatively automatic way and the advantage for these types of stimuli might be important for social interaction (Friesen & Kingstone, 1998; Ohman, Flykt, & Esteves, 2001). The current research is not focussed on social cognition and so none of the experimental pictures featured people, animals or items with any particularly strong emotional value.

2.3 Saliency map model

In order to manipulate or make conclusions about the visual saliency of objects or scene regions, the experimental pictures were screened using a computational implementation of the saliency map model. While other similar models are available (see section 1.5) Itti and Koch's (2000) model was used, as it is particularly explicit and testable. The model is described in Itti and Koch (2000) and here in section 1.4. Since its proposal the Itti and Koch model has been refined and updated, although for the most part the core assumptions (and predictions) of the model remain the same. The experiments reported here use a version of the model compiled from source code available at http://ilab.usc and downloaded in May 2004.

A possible criticism of the model is that it has several free parameters and settings that have the potential to change the predictions of the model. These include the size of the "focus of attention" (FOA), how long the inhibition of selected regions lasts and how the feature channels are weighted and normalised before combining into the saliency map. In the current work the standard parameters were used, as far as possible. For example the default setting for the FOA is a size equivalent to $1/16^{th}$ of the image, which it is argued is a realistic estimate of the resolution of human visual attention. It is largely beyond the scope of this thesis to discuss in detail the results of using different modelling parameters, although these are mentioned where relevant. Instead the aim is to evaluate the general validity of an eye guidance system driven by saliency.

The Itti and Koch model takes as its input a colour digital image and performs computations on the features within it to produce a saliency map and a prediction of the points which should be selected by saliency-based covert and overt attention (see Figure 1.1 in previous chapter). This output is available in several forms. Firstly, feature maps and conspicuity maps from various stages, as well as the final saliency map, are available as two-dimensional arrays of values. The value of each pixel in these maps corresponds to the feature strength at this point in the image. These maps are used for some analysis in subsequent chapters. However the saliency map model is not theory-neutral with regard to the mechanics of attentional selection: attention and eye movements are controlled by a winner take all network with inhibition of return. As a result the experiments here also use the predicted eye movements, or the saliency "rank", as a measure of the saliency of points in the scene. The sequence of points selected can be output in terms of spatial coordinates. So, in Experiment 1 for example, target objects are classified as medium or low saliency based on the number of simulated shifts it takes the model to select the object's location.

Finally, it is worth emphasising that visual saliency measured in this way is relative to the other items in the scene. Thus, the same object can be made more or less salient by placing it in an emptier or more cluttered scene respectively.

2.4 Eye tracking apparatus and methodology

This thesis uses eye movements as an index of attention in natural scenes. Individuals are able to make covert shifts of attention, which are implied from facilitated processing at certain spatial locations whilst maintaining fixation or at intervals too short to permit eye movements (Posner, 1980). Visual search paradigms often explore attention in this way. However, when permitted, humans move their eyes in order to direct attention, and these movements reflect the search process. To investigate the allocation of attention in a naturalistic task it is therefore most appropriate to look at eye movements.

2.4.1 Apparatus

All the experiments reported record eye movements using a video-based eye tracker, and the general set-up will be described here. In each case the EyeLink system was used (SR Research, Mississauga, Canada). This system monitors eye position using infra-red cameras to detect the position of the pupil in one or both eyes, and optionally by tracking the corneal reflection as well. With the exception of experiments conducted with a patient (see Chapter 12) the head-mounted EyeLink systems (EyeLink I and II) were used. In these systems the eye cameras are mounted on a headset and the participants are relatively unconstrained. The system tracks and compensates for small head movements using LED sensors on the screen. In order to minimise these movements a chin rest was used throughout and participants were told to keep as still as possible during experimental trials. The main difference between the EyeLink I and EyeLink II models is the sampling rate; they record at 250 Hz and 500 Hz respectively. The EyeLink 1000 system used in Chapter 12 works in exactly the same way but the eye cameras are mounted on the desk, in front and below the observer. For this set up a full head frame was used to reduce head movements. The EyeLink 1000 sampled pupil position at 1000 Hz.

Two linked computers supported the experimental procedure. The first controlled the presentation of experimental stimuli on a CRT monitor positioned in front of the participant. With the EyeLink I experiments a monitor with a screen of dimensions 34 x 27cm was used, whilst with the EyeLink II and EyeLink 1000 the screen was 40 x 30cm. The distance from the observer to the screen was kept constant by using a chin rest, and was chosen in each case to provide accurate tracking with a large image. The

distance in each case was 60cm and the resulting size of display in degrees of visual angle was 31˚ x 25˚ for the EyeLink I system, whilst for the EyeLink II and EyeLink 1000 set-ups the display was slightly larger (34˚ x 27˚). The second computer processed and stored the eye movement data, which included messages indicating exact timings of when stimuli appeared on the participant's screen. Responses were entered using either a keyboard or a game pad, which was moved close to the participants hands in order to avoid loss of data from the participant looking down at the controls.

2.4.2 Calibration procedure

The calibration procedure was the same for all eye-tracking systems. Prior to each recording participants were shown a series of 9 dots one at a time on the monitor. These locations formed a 3 by 3 grid evenly spaced across the screen and allowed the system to extrapolate gaze location from the pupil image. The experimenter confirmed steady fixation on each point and this was repeated when necessary. A validation procedure then measured the mean error in this calculation and if this was greater than 0.5˚ calibration was repeated. In those cases where a good calibration could not be attained, the participant was replaced.

Due to small amounts of "drift" which occur in the system (for example due to slight head movements or headband slippage), it is desirable to check the calibration regularly. In most experiments this was accomplished by a "drift correct" procedure: a single fixation point was presented (normally in the centre of the screen) in between every trial. Stable fixation on this point was confirmed and this realigned the calibration. If the offset between the actual target position and the computed eye position was larger than 5˚, or if a stable fixation was not detected, then calibration was repeated.

2.4.3 Data analysis

The EyeLink systems recorded sample data indicating the location of gaze, in pixel coordinates, every few milliseconds. Before any analysis was carried out, these samples were parsed into fixation and saccade events (and blinks) using the EyeLink software. This event parser identifies epochs in the data file where a saccade is occurring by calculating the distance between gaze position in different samples and implementing

motion, velocity and acceleration thresholds. For the EyeLink I the standard 'cognitive' thresholds were used which were 0.15˚, 30˚/s and 4000˚/s^2 respectively. For the EyeLink II and the EyeLink 1000 these thresholds were 0.1˚, 30˚/s and 8000˚/ s^2. The parsing algorithm identified a saccade when the velocity or the acceleration computed from consecutive samples exceeded the relevant threshold. Saccade onset time was delayed until eye position had moved as far as the motion threshold. Fixations were identified as the epochs when a saccade was not occurring. Similar thresholds are commonly used in scene perception research and they are designed to reduce the numbers of very small or very fast saccades. The estimates of fixation duration and saccadic amplitude may be slightly greater than if a more sensitive algorithm were used.

Various different measures were derived from the eye events during the experiment, and these are described in each experiment. The measures were straightforwardly computed using macros written in Visual Basic (EyeLink I) or the EyeLink DataViewer software. The eye movement measures will be discussed in greater depth later but it is worth reviewing the most common ones here. These are in addition to manual response measures (accuracy and reaction time) that were calculated from the keyboard responses logged in the data file.

The location of a fixation gives an indication of the part of an image or display that is being processed. Where particular regions or objects were of interest their coordinates were defined and the fixations that landed within them were labelled. The time, and the number of fixations elsewhere, before fixating a region are measures of the potency of that region in attracting attention. When viewing time is limited the proportion of trials where a region is fixated at least once (or the probability of inspection), and the number and proportion of fixations which land in the region also measure how far the eye guidance system is affected by this object or feature. Thus looking at the regions that are fixated more often and earlier in a trial can give information about which features guide eye movements. In visual search the time to fixate the target is a measure of the efficiency of search at excluding distractors and selecting the task-relevant objects. The duration of fixations is usually thought to reflect the extent or difficulty of the cognitive processing occurring (Rayner, 1998). The first

fixation duration on a region therefore gives information about the processing performed once the eyes have moved there. In some cases it is more informative to sum the duration of all fixations made on a region before moving beyond its bounds. This measure can be defined as the gaze duration on this region.

In terms of saccades, the latency and amplitude of the eye movements made can also give important information about the dynamics of overt attention. Saccade latency is the time taken to move the eyes. Reflexive shifts to more potent attractors of attention might be expected to lead to saccades with a lower latency. Saccadic amplitude indicates the size of the attention shift and so can give some information about the efficiency of search and the degree to which the saccade has been targeted using peripheral information. In Chapter 11 saccade direction is also explored in order to examine the distribution of eye movements in natural scene viewing.

These are just a few of the many different measures that are possible in eye-tracking research. The implications of these measures are discussed in subsequent chapters. In general, eye movement research tends to look at eye movement events individually. Chapters 4 and 5 explore some of the issues arising when sequences of fixations – *scanpaths* or *scan patterns* - are analysed. All measures in this thesis, once derived, were averaged across trials, and subject means were analysed statistically. In most cases, means were compared between conditions using analysis of variance (ANOVA). An alpha level of .05 was used for all tests unless otherwise stated. Main effects across multiple levels were compared using post hoc, pairwise comparisons. Tables giving full details of all the statistics quoted are included in Appendix D, where they are organised by experiment.

3 Saliency and task

3.1 Introduction

This chapter looks at the potency of objects in natural scenes to attract overt attention as a function of their visual saliency, as coded by the saliency map, and also as a function of the task being undertaken. There are two main reasons why the interaction of task and saliency is of interest. Firstly, consideration of task may resolve questions about how far overt attentional selection is bottom-up. At least part of the variation in estimating the stimulus dependence of attention may be due to failing to take (implicit or explicit) task demands into account. Vision is an active process and even when "free viewing" scenes it is likely that participants' knowledge and presumptions of what behaviour is expected of them will influence their performance.

Secondly, it may allow conclusions about exactly how top-down and bottom-up processes interact. The present study compares three tasks: a "memory encoding" task which requires participants to inspect photographs in preparation for a recognition test, encouraging them to look at the details and two search tasks. Two search variants are used, one in which the target is defined by a broad category ("category search"), and one in which it is a specific instance of the same category ("instance search").

In all tasks the attention given to objects with known saliency will be measured. If task demands influence a visual saliency map by selectively weighting those features present in preconceived targets, a number of predictions can be made. Fixations in a memory task should be guided more on the basis of visual salience than those in a search task as in the former there is no specific target. As a result there will be little or no prior knowledge about specific informative features to weight a saliency map and fixations should be determined by default, bottom-up control. Top-down guidance in this task should be less pronounced than in search, where a clear target could bias the saliency map. In category and instance search the same objects, with constant bottom-up saliency, will function as targets. Previous studies of search have assumed that detailed iconic representations of the target are available, although in the real world this may not be the case. Some researchers have suggested that search is more efficient when a more exact representation of the target is given in advance (Schmidt & Zelinsky, 2007; Vickery,

King, & Jiang, 2005; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004). In instance search more feature information about the target is available than in category search. Does such information make search more efficient and less likely to be distracted by other salient areas? If top-down instructions can be input into the same saliency map used in the memory conditions, in the form of filtering or weighting by expected features, then this process might well be enhanced in instance search, due to more specific target information. This prediction relies on some assumptions and so it is a tentative one. It is assumed that a target indicated by a verbal label can be efficiently translated into a set of features, that this set will be more restricted in instance search, and that this difference will be exploited by the eye guidance system.

Similar assumptions are present in recent models of search (Navakpakkam & Itti, 2005; Rao et al, 2002). It should also be noted that if the search process is based on a bottom-up saliency map then some features not present in the target might still be salient enough to attract fixations, even after target-based saliency reduction. Does raw visual salience have less impact in instance search than in category search (as there is more specific target information to bias the model)? In other words, is cognitive override by task more frequent in instance search?

3.2 Experiment 1(a): Object saliency in memory encoding and search

3.2.1 Method

Participants

Three groups, each containing 15 student volunteers participated in this experiment. All were naive to the purpose of the experiment. Inclusion in the study was contingent on reliable eye tracking calibration and in particular on maintaining a central fixation at the beginning of the majority of trials. Three participants were replaced, as they did not meet these criteria.

Stimuli, design and apparatus

The same set of 48 digital photographs was used for stimuli in each condition, and these were taken using a 5MP digital camera. All photographs showed office scenes. Many

different instances were used and all the scenes contained standard office furniture (desk, chairs, computer) along with a selection of smaller office objects such as books and stationery. Pictures contained similar amounts of office clutter, and no scene was used more than once. There were 24 experimental stimuli, all of which contained a principal object (a piece of fruit) alongside the other items. Four types of fruit were used (apple, lemon, orange and pear), with six pictures containing each type. These objects were chosen as they are of a similar size and have smooth contours and constant colouring, factors important in determining visual saliency in the saliency model. In the interests of clarity, the fruit will be referred to as targets, although they were only highlighted as such to participants in two of the three task conditions. In each picture, the target could be located anywhere in the scene (within physical scene constraints) but was positioned at a distance of 12˚ from the centre. The visual saliency of the target was manipulated as described below.

A saliency map was computed for each picture that allowed the relative saliency of different regions to be measured and used as a selection criterion. The saliency map was generated using the Itti and Koch (2000) model discussed previously. An example is included in Figure 3.1 (bottom panels). The present experiment is principally concerned with the order of fixation and so the rank of the regions selected by the winner-take-all network was used as selection criteria rather than specific values. The choice of stimuli was made on the basis that none of the target objects were highly salient enough to feature in the first five peaks predicted by the model. Lower saliency objects were chosen here in order to extend scrutiny of the saliency map model to longer and more natural viewing periods, as opposed to using the number one most salient object in the scene, as was the case in Underwood, Foulsham, van Loon, Humphreys and Bloyce (2006). Targets were further classified as medium saliency (featuring between the 5th and the 10th peak) or low saliency (featuring after the 10th peak), allowing the effect of saliency to be explored. An equal number of medium and low saliency pictures were included, with each target fruit equally represented in both. In practice, the saliency measure was an indicator of how much the target stood out from its background compared to the other distinctive objects in the scene. Thus the same object could be made less salient by

placing it on a background of similar brightness and colour. Target objects were un-occluded. As a further control, the most salient region in the picture (the first predicted peak) was always on the opposite side of the picture from the target. Figure 3.1 shows an example stimulus with graphical output from the saliency software indicating the salient regions in terms of their predicted ordinal saliency. A further 24 pictures did not contain a target and were used as controls. In addition, three sets of 8 practice pictures were prepared to familiarise subjects with the tasks.

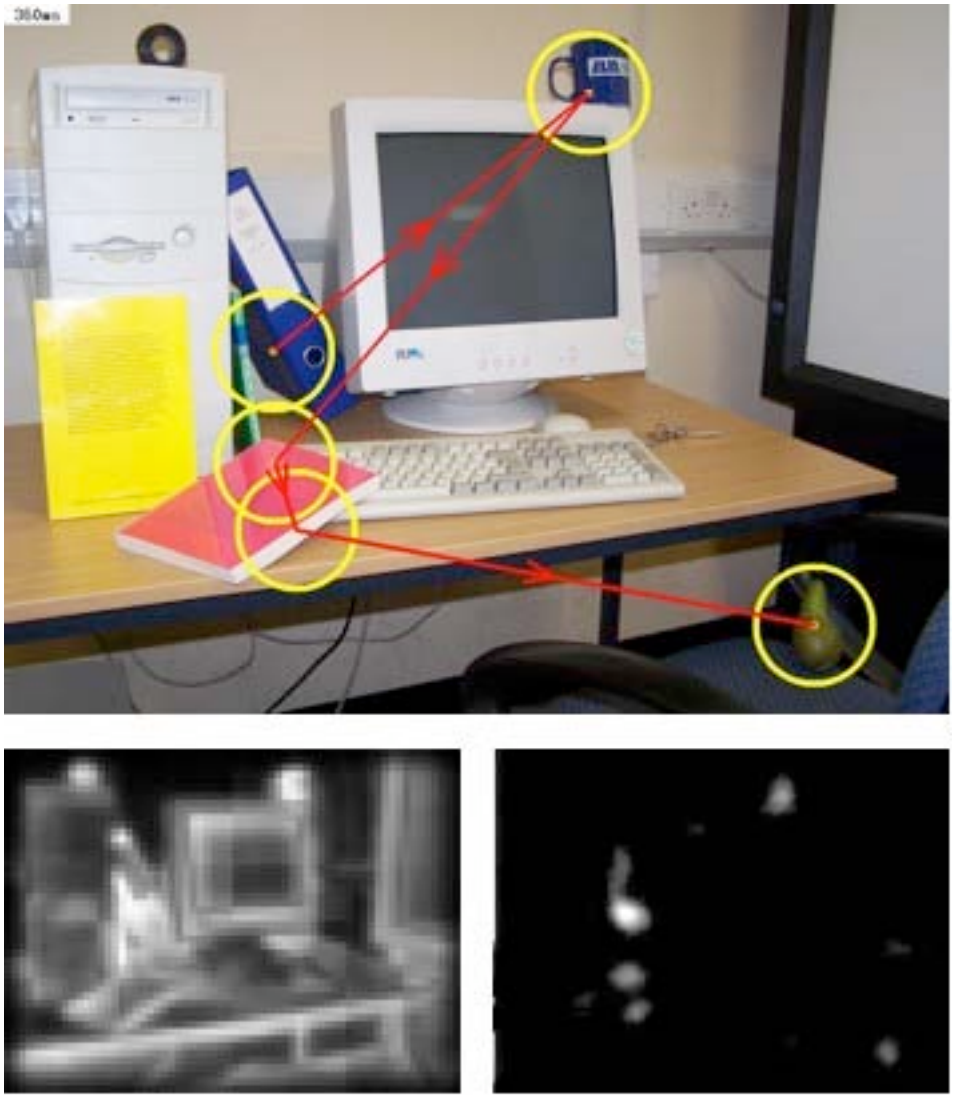


Figure 3.1. An example stimulus from the medium saliency condition, with ranks from the saliency model (top). The area inside the circles indicates the focus of attention which,, among other things, determines the region of inhibition following a fixation. A non-normalised saliency map (bottom left) which is formed from the combination of intensity, colour and orientation conspicuity maps, is included with the final saliency map (bottom right), formed after normalisation and lateral competition processes. In both cases brighter areas indicate higher saliency. Note that while the corner of a folder and the mug feature early in the saliency map, the target pear is ranked 5th.

Visual saliency of the target was a within-subjects manipulation with two levels (low vs. medium). In addition, subjects were randomly allocated to one of three task conditions. These were a task simulating encoding for a memory test ("memory encoding") and two search tasks ("category search" and "instance search"). Thus the between-groups factor of task had three independent levels.

The EyeLink I eye tracker was used to record eye movements, with the standard set-up described in Chapter 2. Responses were entered using a keyboard.

Procedure

A preliminary calibration phase ensured that the apparatus was recording correctly. Participants were then shown written instructions on the screen. Prior to each picture, a drift correct marker and then a fixation cross, both in the centre of the screen, were presented which confirmed that initial fixation was in the centre.

In the memory encoding condition, instructions told the participant to view the scenes "in preparation for a memory test". Viewing was self-paced, and subjects were told to press a key to see the next picture. In a short training phase, subjects were shown some practice pictures followed by a two-alternative forced-choice recognition test featuring one picture they had seen previously and one that differed in the location or presence of an object (neither contained a target from the main experiment). This memory test was not given other than in the practice session, although most subjects expected it, as our concern was with attention and eye guidance during scene perception. This task was designed to simulate viewing of the whole scene with no particular preference for any one object, and has been used previously by Henderson et al. (1999, Experiment 1) and Underwood et al (2006). Following the practice session, all 48 pictures were shown in a randomised order with each one being terminated by the participant's key press. The target will be labelled as such in order to conform to the terminology of the other conditions, although in this condition there was no reason to look at the object and participants had no knowledge of its significance. Few subjects identified any significance of the targets in this condition due to the large number of control pictures without fruit.

In the first search condition, instructions informed the participant that the task was to search for the target, a piece of fruit, in each picture. This task was therefore a category search, looking for members of the category "fruit". If the target was present, participants had to press the "Y' key, and otherwise the "N" key, as quickly and accurately as possible. Following a practice phase, all 48 pictures (half of which contained the target) were presented in a randomised order, with stimulus offset triggered by the response.

The second search task was similar, but here the target was a particular instance of the category "fruit" (apple, lemon, orange or pear). This target was indicated by written instructions at the beginning of each of four blocks (one for each type of fruit, e.g. "the target for this block is an APPLE"). In each case the subjects had to respond by pressing the "Y" or "N' key to indicate if the target was present. Each block consisted of 12 pictures in a randomised order, six of which contained the target (three of low and three of medium saliency) and six of which contained no target. All participants viewed all four blocks in a random order. This condition will be referred to as "instance search".

3.2.2 Results

A range of eye movement measures was computed from the raw data that showed fixation coordinates for each time sample. Although targets varied slightly in size (with mean dimensions of 1.9° by 2.1°), target fixations were identified where fixation coordinates were within a standard 100 pixels (3.1°) square that was centred on the target and which fitted all instances. Fixations were excluded if less than 100ms in duration. In addition, fixation location at picture onset had to lie within one degree of the centre of the screen for that trial to be included. This was encouraged by the central fixation cross prior to each picture and was used as a strict way of ensuring the eccentricity of the target. This condition was not met for 14% of all trials and in these cases no further measures for that trial were included. Trials in the search tasks that led to an incorrect response were also removed. Figure 3.2 depicts an example of the scan patterns made by observers on each of the three tasks whilst viewing the same stimulus as that in Figure 3.1.

Figure 3.2. Example scan patterns for one participant during memory encoding (top), category search (middle) and instance search (bottom). The first fixation, which was necessarily within one degree of the centre, is shown by a square and subsequent fixations by circles. Shape size is proportional to fixation duration. Lines indicate saccades, with arrows representing direction. In the memory encoding example, the target (a pear) was fixated on the 19th fixation (eye movements after this are omitted). In the search examples, the target is fixated on the 4th and 3rd fixations for category and instance search respectively.

The measures taken reflect the hypothesis that visual saliency will affect attentional selection, and therefore how soon and how often an object is fixated, and maybe other cognitive processing, which might be indicated by how long objects and scenes are inspected. Finally, for the search tasks, the proportion of correct responses to target pictures (that is, responding "Y") was analysed. In each case, mean values were calculated across participants for each saliency level and task condition. A summary of the measures taken is provided in Table 3.1.

Task		**Memory encoding**		**Category search**		**Instance search**	
Saliency		**Low**	**Medium**	**Low**	**Medium**	**Low**	**Medium**
Ordinal fixation on target or end of trial	***M***	16.51	13.59	4.26	3.61	4.19	3.76
	SEM	*1.70*	*1.07*	*0.23*	*0.14*	*0.26*	*0.21*
Probability of target being fixated	***M***	0.74	0.83	0.89	0.91	0.91	0.88
	SEM	*0.05*	*0.04*	*0.05*	*0.06*	*0.03*	*0.04*
First-gaze duration (ms)	***M***	581	658	432	427	524	567
	SEM	*66*	*58*	*23*	*25*	*65*	*70*
Total inspection duration (ms)	***M***	9210	9044	1275	1188	1457	1355
	SEM	*1189*	*1204*	*64*	*53*	*107*	*99*
Number of discrete target fixations	***M***	2.20	2.56	1.46	1.46	1.30	1.63
	SEM	*0.31*	*0.31*	*0.10*	*0.12*	*0.10*	*0.17*
Proportion of correct responses	***M***			0.84	0.87	0.85	0.91
	SEM			*0.05*	*0.03*	*0.06*	*0.03*

Table 3.1. Means and standard errors for all the measures taken, organised by task condition and the visual saliency of the target.

A series of ANOVA tests were performed to determine statistical reliability. All pairwise comparisons were post hoc Scheffé tests. Each measure will be discussed in turn.

Ordinal fixation on target or end of trial

The number of fixations on the picture leading up to fixation of a target is an indicator of how quickly that object attracts attention. Targets that are potent in attracting attention will be fixated after fewer fixations than other objects. Targets that are less potent will be fixated after more fixations elsewhere in the scene, or the trial will be terminated before target fixation. This measure was therefore analysed to explore whether medium targets attracted attention earlier than low targets, irrespective of task. The earliest a target could be fixated was on the second fixation, as the first was necessarily in the centre of the display. The highest value this measure could have was the total number of fixations on the picture which varied (viewing was self-paced) but had a mean of 30.5, 4.6 and 4.7 in the memory, category search and instance search conditions (see analysis of total inspection duration which reflects this measure). Figure 3.3 displays the cumulative probability of fixation for each fixation since picture onset, and for each separate task condition.

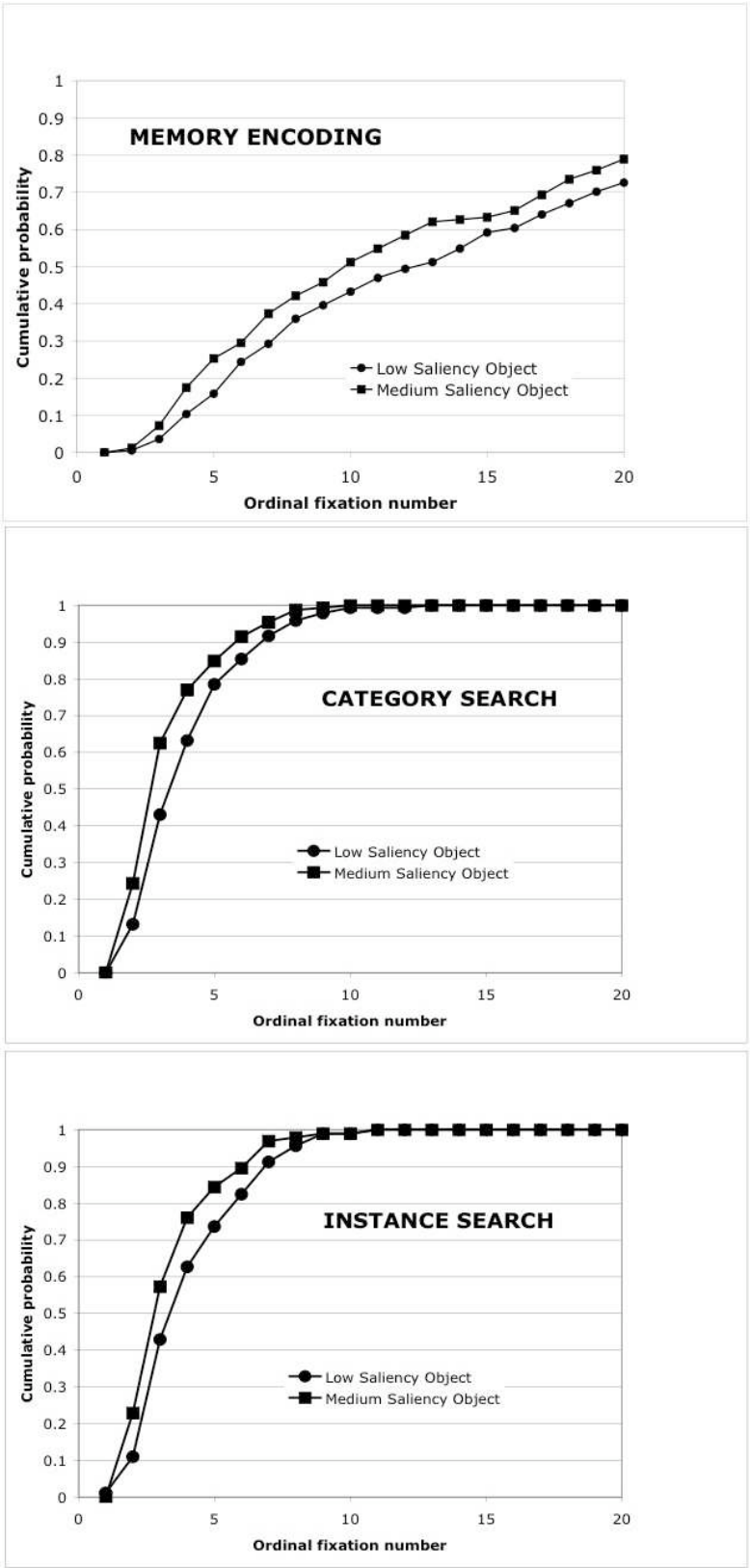


Figure 3.3. The cumulative probability of a target being fixated at least once as a function of ordinal fixation number since display onset for each task condition: memory encoding (top) category search (middle) and instance search (bottom). Targets were much more likely to be fixated earlier in the search tasks than in the memory task. Values may differ from those elsewhere as they do not include those trials where the target was not fixated.

A two-way ANOVA with the within-subjects factor of visual saliency and the between-groups factor of task was carried out on the participant means. The results showed a highly significant effect of saliency, $F(1,42) = 13.56$, $MSE = 2.97$, $p = .001$, with medium targets fixated before low targets (ordinal fixations, Ms = 6.99 and 8.32 respectively). As would be expected, task had a very reliable effect, $F(2,42) = 67.94$, $MSE = 18.12$, $p < .001$, and pairwise comparisons revealed that targets were fixated later when participants viewed pictures for a memory test (where there was no task requirement to look at the target, $M = 15.05$) than in either category search ($M = 3.93$) or instance search ($M = 3.98$), both ps $< .001$. The two search conditions were not significantly different.

There was an interesting interaction between task and saliency, $F(2,42) = 4.81$, $MSE = 2.97$, $p = .013$. Simple main effects analysis revealed that while saliency had an effect in memory encoding $F(1,42) = 21.63$, $MSE = 2.965$, $p < .001$, in all other cases it was not significant. Task had a reliable effect on both levels of saliency (low, $F(2,84) = 71.0$, $MSE = 10.54$, $p < .0001$; medium, $F(2,84) = 46.0$, $MSE = 10.54$, $p < .0001$).

Probability of target fixation

This measure was taken as a second indicator of the potency of the target in capturing attention. It was calculated from the proportion of trials where fixation lay within the target region at least once during stimulus presentation. If saliency is important in all tasks then fixations will be more likely to lie on medium targets than low targets.

As with the previous measure, a two-way ANOVA was performed on the participant means and while the main effect of saliency approached significance, $F(1,42) = 3.78$, $MSE = 0.0041$, $p = .059$, the effect of task was not significant, $F(2,42) = 2.07$, $MSE = 0.0629$, $p = .139$. There was however a significant interaction between the two, $F(1,42) = 7.32$, $MSE = 0.0041$, $p = 0.002$. Analysis of simple main effects showed that while task had a significant effect on the probability of fixating low saliency targets, $F(2,84) = 4.047$, $MSE = 0.034$, $p = 0.021$, this was not significant with medium saliency targets. In addition, there was a simple main effect of saliency only when encoding for a memory test, $F(1,42) = 15.99$, $MSE = 0.004$, $p = 0.0003$. This indicated that, in this

condition, medium saliency targets were more likely to be fixated than low saliency targets.

First gaze duration on target

The duration of the first gaze on an object is an index of how difficult processing that object is (Rayner, 1998). Gaze is the sum duration of all consecutive target fixations before fixating outside the region, including the first fixation duration. As the meaning or task demands related to medium and low targets did not differ saliency should not effect gaze duration. This measure also served as a control that targets did not differ in other ways, such as ease of processing once fixated.

There was no significant main effect of saliency, $F(1,42) = 2.50$, $MSE = 12766$, $p = .12$. The first gaze duration was not different for low and medium saliency objects. There was a main effect of task on gaze, however $F(2,42) = 3.58$, $MSE = 76789$, $p = .037$. Post hoc comparisons showed that gazes in the memory encoding condition were significantly longer than those in the category search condition (619ms and 430ms respectively; $p < .05$). The other comparisons were not reliable. There was no significant interaction of saliency and task on first gaze durations $F(2,42) < 1$.

Total picture inspection duration

This measure was taken as the interval between picture onset and the terminating key press response. As such it was an indicator of the time required to perform the task before moving on. While there was a highly significant effect of task, $F(2,42) = 43.14$, $MSE = 14138314$, $p < .001$, neither the within-subjects factor of saliency, $F(1,42) < 1$, nor the interaction, $F(2,42) < 1$, reached significance. As might be expected, comparisons between the different task conditions revealed that pictures were inspected for much longer in memory encoding ($M = 9127$ms), where the task was more challenging and there was no target end-point, than in either category search ($M = 1231$ms) or instance search ($M = 1406$ms; both ps $< .001$). There was no significant difference between the total picture inspection duration in the two search conditions.

Number of target fixations per trial

This measure, the number of times a target was separately fixated on any one trial, was taken to investigate whether certain targets were often refixated. The score for trials where the target was not fixated was zero. The ANOVA test gave a significant effect of saliency, $F(1,42) = 8.25$, $MSE = 0.146$, $p = .006$, indicating that medium saliency targets were fixated more times per trial ($M = 1.88$) than low saliency targets ($M = 1.65$). There was also a significant effect of task, $F(2,42) = 7.54$, $MSE = 1.12$, $p = .002$, and post hoc comparisons indicated that targets were fixated more times per trial ($M = 2.38$) in the memory encoding condition than in either search condition (both Ms = 1.46, ps < .01. No other differences were significant. The two factors of task and saliency did not interact, $F(2,42) = 2.05$, $MSE = 0.146$, $p = .141$.

Proportion of correct responses

For the two search tasks, the proportion of correct responses to those pictures with each type of target (responding "Y" to target pictures or the "hit rate") was analysed using a two by two ANOVA with saliency (low vs. medium) and task (instance vs. category search). The hit rate was high in all cases and there were no significant effects: saliency, $F(1,28) = 1.16$, $MSE = 0.024$, $p = .291$; task, $F(1,28) = 0.293$, $MSE = 0.0315$, $p = .593$; saliency by task interaction, $F(1,28) = 0.084$, $MSE = 0.024$, $p = .773$. Several participants were 100% accurate and false alarms in these tasks were relatively rare, occurring 8.1% of the time across conditions.

Potency of most salient region

There was a smaller, non-significant effect of target saliency on ordinal target fixation and target fixation probability in the search tasks than in the memory task. A possible objection to these results is that by the 5th or 10th predicted fixation, saliency values are lower. This forces the model to rank regions that may be only marginally different, and so may lead to an unfair evaluation of its performance. Although a complete model should account for the time span of a whole trial, a further test of saliency is to look at fixations on the most salient region in the picture. This region was defined as that ranked first by the saliency model (it is the corner of a folder in Figure 3.1) and was bounded by

a rectangle of the same size as the targets. Fixations on this region were recorded and the proportion of trials where the most salient region was fixated was analysed using the same two-way ANOVA as previously. Task was highly significant, $F(2,42) = 191.2$, $MSE = 0.0215$, $p < .001$, with the most salient region capturing attention much more often in a memory task ($M = 0.84$) than in either search task (category $M = 0.19$; instance $M = 0.21$; ps $< .001$). Interestingly, there was also an effect of saliency, $F(2,42) = 6.31$, $MSE = 0.0091$, $p < .05$, such that the most salient region was fixated more often when the target was medium saliency ($M = 0.44$) than when it was low saliency (0.39). There was no interaction, $F(2,42) = 1.02$, $MSE = 0.0091$, $p = .369$.

A valid objection to this analysis is that memory trials contained more fixations so that even if fixations were allocated randomly the memory task would be expected to contain more fixations on the most salient region. To resolve this, the above analysis was repeated using only the first five fixations from the memory task. Search trials contained five fixations on average, making the two comparable. Task remained significant, $F(2,42) = 4.00$, $MSE = 0.019$, $p < .05$, indicating that even in the first five fixations the most salient region was more potent at attracting attention in the memory task (0.29) than in the search tasks (means as above, both ps $< .05$). The effect of target saliency remained, $F(2,42) = 5.02$, $MSE = 0.013$, $p < .05$, and the interaction was not significant, $F(2,42) < 1$.

Summary of results

As would be expected the task instructions had a large effect on viewing behaviour with people making more fixations, longer picture inspections and longer first gazes when encoding for a memory test than when searching for something. Search was efficient so that targets were fixated much earlier when subjects were actively looking for them. Saliency had an effect on the ordinal fixation of targets such that medium saliency targets were fixated earlier than low saliency ones, and this was the case even late in the trial.

There were a number of particularly interesting results. Saliency had a significant effect on fixation probability (how often) and ordinal fixation number (how early objects were fixated) only when pictures were viewed for memory encoding, and not during search. In any one trial, targets were refixated more often if they were more

salient. There were no effects of saliency in the two search tasks, and there were no significant differences between category and instance search. Accuracy in the search tasks did not differ due to target saliency or search variant. The most salient region in the scene was more likely to be fixated in the memory task than in the search task, even when differing numbers of fixations were controlled.

3.2.3 Discussion

The experimental manipulation of saliency had a significant effect. Differences in visual saliency caused the objects of interest to be fixated earlier when they were ranked higher according to Itti and Koch's (2000) saliency algorithm. However, this was only the case when viewing in preparation for a memory test, and not when searching for the objects. Higher saliency objects were also more likely to be fixated in the memory task. This suggests that bottom-up selection is important when scanning photographs, but that these effects are not independent of task. The saliency map model correctly predicted which object would be fixated first, as medium saliency targets were on average fixated earlier (than low saliency targets) and were by definition predicted to be fixated earlier. There was a general trend for targets to be fixated later than the ranks generated by the model (for example medium saliency objects were by definition ranked between 5th and 10th by the model but were fixated after 13.6 fixations on average). This suggests that the fit between the model and real data was not perfect. Other models, either modifications of a saliency map taking into account different features or other accounts of bottom-up selection, might do better. In the memory task the objects did not differ semantically in importance with regard to scene or task context but only in low-level discontinuities.

The fact that an effect of saliency can be found fairly late in scene viewing (after more than ten fixations on average) is interesting. Parkhurst et al. (2002) suggested that the effect of saliency decreases over viewing time. If this is the case, the present results indicate that even late in viewing saliency is still a significant factor. Henderson et al.'s (1999) general framework suggests that items are always initially fixated on the basis of saliency, but once acquired can be evaluated in terms of cognitive demands which determine later processing. Henderson et al., along with many other researchers, used line drawings where the saliency discussed here is effectively meaningless, so it is

important that the present research used photographs (as did Underwood et al., 2006) allowing the saliency hypothesis to be tested fully. Saliency had no effect on an index of processing (first gaze duration) suggesting that while bottom-up processes were important for attentional engagement, disengagement was not dependant on this. However, there was a tendency for higher saliency objects to be refixated, despite receiving presumably equal processing on the first gaze. This is shown by the fact that there were on average around two discrete fixations on these targets per trial and this suggests that bottom-up selection continued to be important and may have triggered reflexive shifts when it probably would have been more efficient (in terms of the memorising task) to fixate elsewhere. The dynamics of the saliency map model are not incompatible with this finding as although fixated regions are inhibited this inhibition is transient and may not lead to complete suppression. How strong and for how long is this suppression? The importance of saliency following the first acquisition of a region is worthy of further study.

The scale of the effects of task makes it clear how important this factor is in eye-guidance. When searching for a target as opposed to viewing scenes for a later memory test, scenes were inspected for much less time and targets were fixated much earlier, gazed at for less time and refixated less often. It is problematic to assume that there is any such thing as "free" viewing and as models become more sophisticated it is important that experiments describe and control task conditions carefully, so as to both anchor results in the real world and enhance model predictions. Changes in fixation behaviour with search tasks similar to those found here were reported by Henderson et al. (1999) and Underwood et al. (2006). The results of the present study emphasise that a strong version of the saliency map hypothesis must be rejected on the grounds that cognitive demands can influence eye guidance and that this can happen before the object is first acquired. Comparison between the memory encoding task and the search tasks here shows that search instructions can override visual saliency allowing earlier target fixation. This behaviour might be modelled using modified feature weights based on the target, as in the approach of Navalpakkam and Itti (2005). The difference between medium and low salience targets was reduced to the point where it was not reliable under search

instructions, suggesting that saliency is not important in search. Given the argument that top-down influences take longer to influence viewing, the analysis of fixations on the most salient region is interesting. Despite limiting this to early viewing this region was more often fixated in the memory task than in the search task. The fact that this region was also more likely to be fixated when the target was more salient is hard to explain and may merit further study.

How does cognitive override of attention work? Several researchers have identified the types of top-down knowledge that may be available in a task and the way in which this might interact with bottom-up saliency. In Findlay and Walker's (1999) model the "where" pathway, which determines which regions are fixated, can be affected top-down in three ways. Spatial selection occurs when areas of the salience map are inhibited or potentiated based on knowledge of target locations. Similarly, Torralba (2003) includes location probability as one of several contextual priors which influence attention. Object location predictability was minimised here, and targets were not strongly cued by the gist of the scene, but location bias may have encouraged saccades to some areas of the display (as targets were always the same distance from the centre). Findlay and Walker's (1999) search selection, which has a parallel in Torralba's (2003) target driven control parameter, enhances the saliency of features present in the target. In a similar way, Navalpakkam and Itti's (2005) model weights the saliency map based on learned features of the target. Findlay and Walker's (1999) final process is the slightly underspecified concept of intrinsic salience, which allows some features (such as contours) to be intrinsically potent at capturing attention and some to develop this salience following medium-term learning. The present results might be explained by any one of these processes in that overt attention was presumably drawn to the targets in the search task on the basis of features. Some more detailed conclusions are possible, however. If search selection proceeds by potentiating a saliency map then saliency should still have an effect on the time to fixate the target (both medium and low saliency target regions will be potentiated as they contain target features, but medium saliency targets will still produce a higher peak). There was no significant effect of saliency in the search conditions here, suggesting that this conceptualisation of the effect of task may be

incorrect in this case. It was not the case that there was a difference in the influence of visual saliency between the two search task variants. This might be predicted as in instance search more specific information about target features is available to bias the map. For example, in the category search while likely shape and scale features might be primed (fruit targets are round and of a similar size), in the instance search, colour was also identified by the target description (lemon targets are yellow, not green or orange like pears and oranges). In this case, there would be fewer possible saccade targets so search might be expected to be more efficient and less susceptible to interference from bottom-up visual salience of other regions. Vickery et al. (2005) reported that visual search is quicker when an exact representation of the target is known than a general category, but the total inspection durations in the two search tasks here (which was also the time to respond) does not show this trend. There were no significant differences in the measures obtained between the two types of search, so if more information about the target was available, it does not appear to have been used in moving the eyes and responding more efficiently.

Williams, Henderson & Zacks (2005) report that memory for (and attention to) distracters in visual search was greater for objects which shared target features. While no distracters were specified here, predictions can be made for eye movements to distracters of various types in an instance search of natural scenes. Semantic category distracters (for example a banana while searching for an apple) and featural distracters (for example a green ball while searching for an apple) should attract attention more than unrelated objects in the scene, and might do so to different degrees. If saliency is unimportant in search, as the results presented here suggest, the saliency of such distracters will not affect their ability to draw attention.

3.3 Experiment 1(b): Speeded category search

3.3.1 Introduction

In Experiment 1(a) the search task was relatively easy; participants had as long as they needed to perform the task and were highly accurate. This raises the possibility that a ceiling effect in some of the measures may have obscured an effect of saliency in category and/or instance search. In other words, people may have been so good at finding the target that the difference between medium and low salient trials was non-significant, when there was in fact a difference. Although participants were instructed to respond as quickly as possible, the absence of a time constraint may also have made it less likely that a saliency effect would emerge. It has been suggested that saliency has its effects early in scene viewing (Parkhurst et al., 2002). Van Zoest, Donk and Theeuwes (2004) looked at attentional capture by bottom-up singletons in a simple visual search and suggested that bottom-up effects decay very quickly, being potent only within the first 200-500 ms, around the time when the first saccade is being planned in the natural search task of Experiment 1(a). To address these issues, a supplementary search experiment was carried out with a short time limit in which participants had to respond. Will greater time pressure lead to a greater difference between the accuracy and eye movements made towards medium and low saliency targets than in Experiment 1(a)?

3.3.2 Method

A new group of 15 student volunteers took part in Experiment 1(b). The stimuli and design were exactly the same as those in the category search condition of Experiment 1(a). The EyeLink II system was used, giving a higher temporal resolution than previously.

The procedure in this experiment was the same as the category search described above; following a short practice participants were told to search for a piece of fruit and indicate its presence or absence as quickly as possible. Category search was chosen as there were few differences between category and instance in the previous experiment, and it was thought that category search was slightly more difficult. However, in this

experiment a time limit of 500 ms was imposed and feedback was given. This duration is enough for around 2 or 3 saccades and fixations of average length, and pilot sessions suggested it was suitable for above-chance performance. After this time limit had elapsed the scene was offset and a blank screen prompted the participant for a response. Following a response, text on a feedback screen indicated whether it was correct or incorrect.

3.3.3 Results

Manual responses

Accuracy under timed conditions was slightly worse than the category search condition in Experiment 1(a) with mean hit rates of 63% and 71% for low and medium saliency targets respectively, though there were few false alarms (7%). A between groups, one-way ANOVA collapsed across target saliency confirmed that performance in the timed experiment was significantly worse than the category search condition in Experiment 1(a), $F(1,28) = 18.1$, $MSE = 0.014$, $p < .001$. Under time pressure there was no reliable effect of target saliency on hit rate, $F(1,14) = 2.15$, $MSE = 0.02$, $p = .164$. Response time was not analysed as it was thought to reflect both search efficiency in the limited picture exposure and "thinking time" on seeing the response screen.

Eye movement measures

Several eye movement measures were derived to indicate how often and how early the target was acquired. The fixed time limit meant that the total inspection time and the number of fixations and refixations were constrained so these will not be examined. The mean probability of fixating low and medium saliency targets was 0.31 ($SEM = 0.027$) and 0.43 ($SEM = 0.036$) respectively. These are much lower than in Experiment 1, reflecting the increased difficulty under timed conditions. The target was fixated in a greater proportion of trials when it was medium saliency than when it was low saliency, $F(1,14) = 14.06$, $MSE = 0.008$, $p < .005$.

Due to the brief trial duration and the higher temporal resolution of the eye tracker the time until first fixating the target was taken as a measure of how quickly it was reached rather than the ordinal fixation number. Looking only at trials where the target

was fixated the difference between the first fixation time for low (M = 382 ms, SEM = 11.3) and medium (M = 358, SEM = 8.7) targets was not reliable, $F(1,14) = 3.06$, MSE = 1326, $p = .102$. As the target was fixated on less than half the trials (unsurprisingly, given the time constraint and the size of the target) this measure excludes rather a lot of data. An alternative measure of how quickly the target affects gaze allocation is to look at the very first saccade—how close to the target did this land in cases of different saliency? The Euclidian distance between the landing point of the first saccade and the centre of the target region was calculated. The mean distance for low and medium saliency targets was 11.15˚ (SEM = 0.68) and 8.10˚ (SEM = 0.625) respectively, and this difference was reliable, $F(1,14) = 15.50$, $MSE = 4.50$, $p = .001$. After the first saccade people were, on average, closer to medium than low saliency targets.

Finally, the probability of fixating the most salient region was also calculated to see if the difficulty of the speeded task resulted in more fixations on the most salient region. In Experiment 1(b) the most salient region was fixated in .189 of low target trials and .116 of medium target trials (SEM = .029 and .024 respectively). These probabilities were not reliably different, $F(1,14) = 3.25$, $MSE = 0.012$, $p = .093$. When collapsed across conditions the two experiments were not different, $F(1,28) = 2.11$, $MSE = 0.006$, $p = .158$; the time pressure did not lead to the most salient region being fixated any more or less often.

3.3.4 Discussion

The addition of a time constraint revealed some effects of saliency in search and so qualifies the findings of Experiment 1(a). The very brief inspection duration clearly made the task more difficult (as indicated by lower accuracy), although it was still performed better than chance. As previously, the hit rate was not affected by saliency. How did the more difficult task affect the eye movements made and the influence of saliency on these?

Targets were fixated on less than half of the trials. Importantly, while the probability of target fixation did not vary in the category search in Experiment 1(a), under time pressure more salient targets were fixated more often. When fixated, medium saliency targets were not necessarily inspected any earlier. However, target saliency did

have an effect on the very first saccade; this landed closer to medium than to low saliency targets.

Why were the saliency results here different from those in the previous category search? It does not appear to be the case that top-down processes completely override any computation of bottom-up saliency, as here there is a search task where the visual conspicuity of the target has an effect. On the other hand, top-down guidance clearly had an influence; targets were fixated sometimes even within 500 ms, or about 2-3 fixations, despite the fact that the saliency map model predicts they should be selected after a minimum of 5 shifts. If it is assumed that top-down, target feature information is combined with bottom-up saliency in some way then the time limit imposed here may have changed the balance of these processes. It might be that top-down information, particularly that gleaned from the initial gist of the scene, takes time to develop and so there is more room for saliency to have an influence. If this were true then it might be predicted that the most salient region would also receive more fixation when a time limit is imposed, but this was not the case.

An alternative formulation is that top-down information is available from the very first fixation, but that the influence of saliency decreases over time. Saccades to the most salient region would still be avoided, as the region is rarely in line with knowledge of the target, but those to the target would be primed by the bottom-up saliency of the region. The requirement to move the eyes and find the target quicker would therefore cause target saliency to have a greater impact in Experiment 1(b). Van Zoest et al. (2004) argued that saliency is most potent early in visual search, supporting this explanation.

3.4 Conclusions and links forward

The results reported here confirm that low-level saliency is important in determining fixation location and order, but only in certain tasks. There was little evidence in Experiment 1(a) that visual saliency was important in eye-guidance in a search situation. Instead, they suggest that cognitive override in search may be an all-or-nothing process that does not rely on the same saliency map as less targeted viewing. Although previous research has tended to focus on what might be referred to as “default” or “intrinsic” saliency, the present results cannot fully reject a version where this property is weighted

by target characteristics. It is clear that a purely salience based model is incapable of explaining the natural complexities of eye movement behaviour and the findings reported here emphasise some of the difficulties of addressing top-down control within such a framework. Experiment 1(b) shows that target saliency may have an effect in search in some conditions, particularly in displays with a brief duration.

The two tasks addressed here will continue to be explored in this thesis. Subsequent chapters look at naturalistic visual search in more detail. When does target saliency affect search, and do non-target, salient distractors produce predictable effects? In Experiment 2 the fixation locations and eye movement sequences made in a memory task are scrutinised in more detail, both at first inspection and when being recognised later. This will allow a more precise specification of the degree to which saliency affects eye movements in memory encoding, and also of the time course of these effects. Before this, Chapter 4 explores some methods for comparing scanpaths, which will be used in subsequent experiments.

4 Methods for comparing scanpaths

4.1 Introduction

How can one best capture the sequences of eye fixations made during the inspection of scenes? Research tends to focus on measures such as fixation counts, fixation (or gaze) duration and saccade amplitude, and I will use all of these at various points in this thesis. What these measurements have in common is that they generally consider each eye movement event (fixation or saccade) independently. This may mean that important information regarding the sequence of such events is missed. For example, the knowledge that larger saccades or longer fixations are directed at a certain part of an image would not identify whether the viewer was consistently moving their eyes in the same way with certain stimuli or whilst performing certain tasks. A significant body of research is concerned with the pattern or sequence of scanning movements that is associated with a particular viewing period or stimulus. These patterns can be referred to as *scanpaths* (though see later in this introduction for some notes about the connotations of this term). This chapter will examine a specific methodological problem which arises when looking at such patterns: how are two scanpaths from separate viewing periods best compared in order to capture the spatial and temporal information which they contain? The degree of similarity of different scanpaths is important to ascertain in order to assess the extent to which eye movements are driven by the same factors across individuals, tasks and stimuli. The remainder of this introduction will illustrate this problem in the context of previous research examining scanpaths, and the following sections will discuss some solutions in more detail.

4.1.1 Scanpaths are representative of stimulus and task

As I have mentioned, the places where people fixate have been related to certain visual features (as in the saliency map approach) but they also vary according to the task being undertaken. However, few models go beyond looking at individual fixations to modelling a complete scanpath. One of the advantages of the saliency map model is that it gives a relatively straightforward way to predict a sequence of fixations, yet most tests

of this model do not go beyond a correlation between individual fixation locations and saliency. Yarbus (1967) highlighted the fact that scanpaths exhibited when viewing a stimulus are quite different if the viewer is given a different task. In Experiment 1(a) the scanpaths seen were very different in the different task conditions, and here I consider ways to quantify these differences by looking at a whole sequence of fixations, rather than taking an overall measure, such as the number of fixations before reaching a target. The study of eye movements in real world tasks such as making tea (Land, Mennie, & Rusted, 1999) makes it clear that the sequence of gazes is strongly tied to behavioural goals. From this "active vision" perspective scanpaths, as opposed to just individual eye events, are crucially important.

The between-subjects variation in scanpaths has led some to label scanpaths as distinctively idiosyncratic, presumably reflecting personal knowledge, experience or viewing strategy (Choi, Mosley, & Stark, 1995; Leonards & Scott-Samuel, 2005). To study these top-down aspects of overt attention it is useful to be able to compare scanpaths across viewings, stimuli and individuals. Before looking at how to accomplish this in more detail, this review will consider a specific theory for which scanpath comparison is particularly important.

4.1.2 Are scanpaths stored in feature memory?

Scanpath theory is an ambitious set of ideas that was originally proposed in two papers by Noton and Stark (Noton & Stark, 1971; Stark & Noton, 1971). The theory describes scanpaths as controlled by internal, cognitive models representing the viewer's expectations of a scene. These models might represent the saccades involved in viewing a picture or scene as a kind of structure or syntax that binds together the features processed at fixation. People impose their interpretation of a scene onto their viewing of it, selecting locations top-down. When viewing the same scene again, as in the test phase of a recognition experiment, a scanpath which has been previously stored might be re-invoked or checked against the external stimulus. The main evidence for scanpath theory came from experiments showing that scanpaths recurred when stimuli were reviewed in a recognition task. In Noton and Stark (1971) this conclusion was reached based on subjective observation of the patterns shown by each subject and there was no

quantification of the similarity between the scanpaths. Other researchers have reported similarity between scanpaths made when viewing a simple checkerboard stimulus and those made when imagining it later (S. A. Brandt & Stark, 1997; Laeng & Teodorescu, 2002). These papers argue that as the scanpaths at viewing are similar to those during imagery, eye movements must be an integral part of an iconic feature memory. As the scanpaths during imagery are made in the absence of any visual stimulus, this is an extreme example of eye movements being guided top-down.

Little empirical support has been found for scanpath theory and there are common observations that it would seem to have trouble explaining. It is not necessary to move ones eyes to encode or recognise a picture, even if eye movements are used at another time. The apparently large amount of variability within the patterns shown by a single person viewing the same stimulus would also seem to make a strong version of scanpath theory untenable (Mannan et al., 1996). As a result, Henderson (2003) cautions against use of the term scanpath due to its association with this theory. This thesis will continue to use the term, in order to remain consistent with previous literature, but this chapter will not examine the specifics of scanpath theory or any other particular approach to the control of eye movements. Instead, it will concern itself with how to quantify the relationships between scanpaths in the hope of clarifying the systematic, idiosyncratic or repetitive aspects of eye movements.

4.2 Methods for comparing scanpaths

In this section, several different algorithms for comparing scanpaths will be reviewed. In order to compare these methods they were implemented in a set of programs written in Java which process raw fixation files showing coordinates for a series of fixations. These programs are available as an internet applet, or to download at http://www.psychology.nottingham.ac.uk/staff/lpxtf/scanpath/scanpaths.html. They allow the user to quickly and easily compare the results of each measure for a given scanpath comparison. Some examples of the Java code for these algorithms are included in Appendix A.

4.2.1 Distance-based methods

Scanpaths are sequences of points in space. As a result it would seem most appropriate to measure the distance between two scanpaths superimposed on the same visual area. A metric developed by Mannan and colleagues (Mannan et al., 1995, 1996; Mannan, Ruddock, & Wooding, 1997) computes the similarity between scanpaths by measuring the distance between each fixation in one set and its nearest neighbour in the other. Scanpaths that are more similar, in the sense that they dwell on locations close to each other, will show a smaller average linear distance. Figure 4.1 depicts this comparison. The average linear distance is defined as *D*, where

$$D^2 = \frac{n_1 \sum_{j=1}^{n_2} d_{2j}^2 + n_2 \sum_{i=1}^{n_1} d_{1i}^2}{2 n_1 n_2 (a^2 + b^2)} \qquad (1)$$

and where n_1 and n_2 are the number of fixations in each scanpath and *a* and *b* are the dimensions of the image. d_{1i} is the distance between the *ith* fixation in the first set and its nearest fixation in the second set, and d_{2j} is the same distance for the *jth* fixation in the second set.

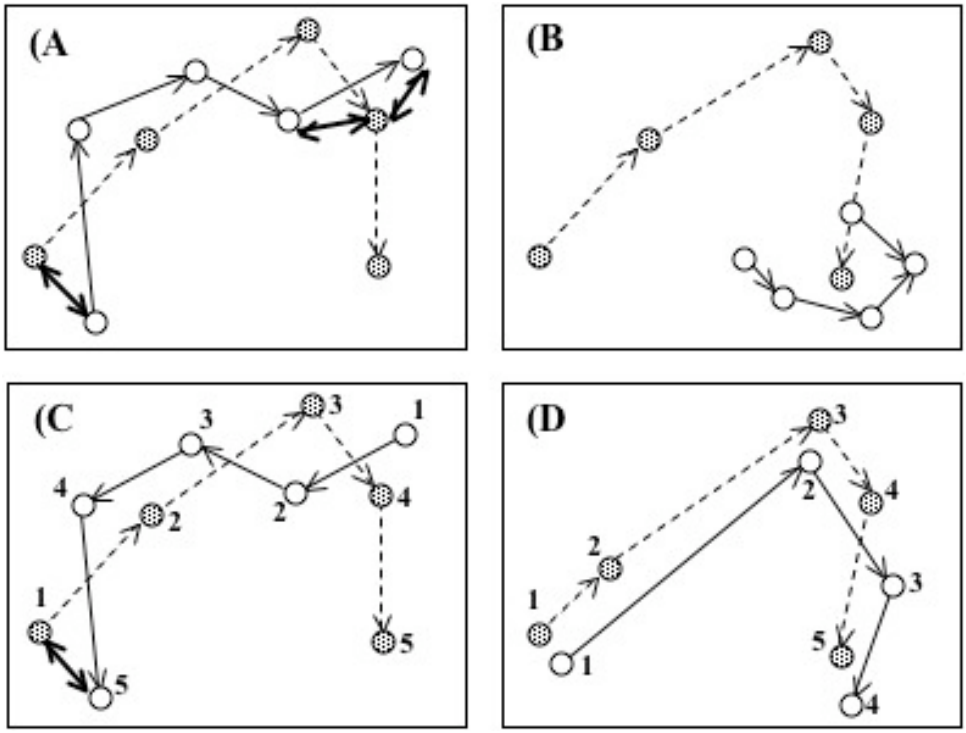


Figure 4.1. Calculating the linear distance between two hypothetical scanpaths. Circles show fixation locations and are linked by lines indicating saccades, with arrows indicating the direction of movement. (A) Each fixation is compared to its closest neighbour in the other scanpath. This distance is illustrated for three fixations (bold arrows). (B) This metric is confounded by differences in the spatial variability of the two scanpaths. All of the fixations in one scanpath (open circles) will be compared to just one in the other set, leading to a low mean distance despite very different patterns. (C) The metric ignores sequence information. Ordinal position is emphasised with numbers. Note that the first fixation is compared to the fifth fixation in the other set. (D) If each fixation is compared with that in the same serial position, small differences skew later comparisons. In the case illustrated, despite broadly similar scanpaths, the distance between second and subsequent fixations is large.

This measure is easy to compute from fixation coordinates. In addition, it is robust for scanpaths with different numbers of fixations and is scaled relative to the size of stimulus being viewed (due to the term $(a_2 + b_2)$). In order to produce an estimate of the absolute degree of similarity, Mannan et al (1995) compute the similarity index, I_s, by comparing the average linear distance between two scanpaths with that between randomly generated scanpaths of the same number of fixations (D_{rand})

$$I_s = \left(1 - \frac{D}{D_{rand}}\right)100 \qquad (2)$$

This gives a value between 0 (chance similarity) and 100 (identical scanpaths), with negative values indicating scanpaths that are more different than expected by chance. Thus this measure includes an estimate of the reliability of the similarity found. Scanpaths that appear similar purely by chance will not differ from the randomly generated scanpaths and will give a similarity index close to zero. Simulating the distance between randomly generated scanpaths (D_{rand}) produces the normally distributed similarity that would be expected from chance or uniform scanning. This distribution is examined by Mannan et al (1995) and illustrated in Figure 4.2. For a constant display size, the mean random distance gets smaller as more fixations are added to the scanpath (as n_1 and n_2, which do not have to be equal, increase). As more and more fixations cover the display area, the likelihood of any one fixation being near to another increases. It is important that comparison metrics take this into account (as the I_s parameter does): longer scanpaths with more fixations will seem more similar by chance.

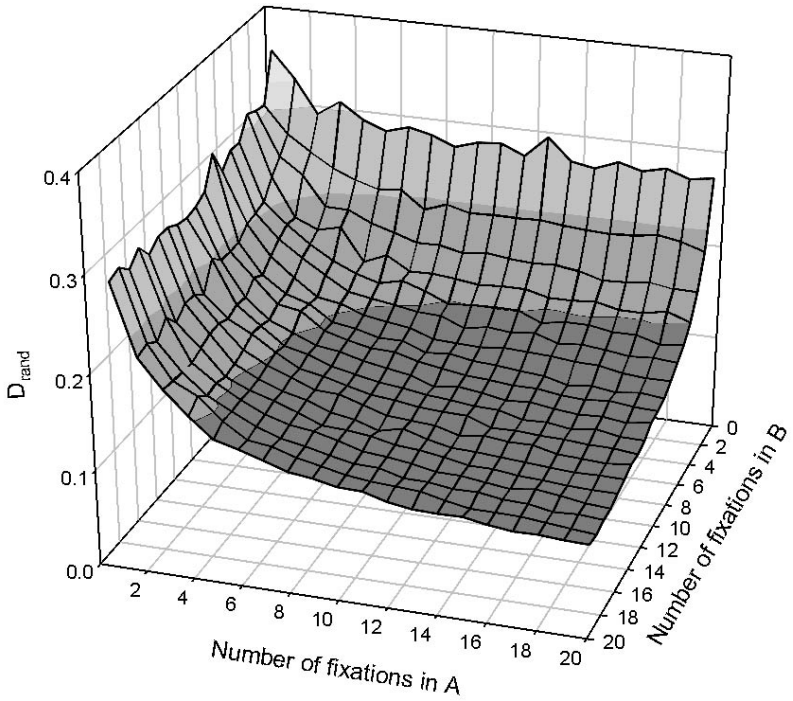


Figure 4.2. The randomly generated, mean linear-distance (D_{rand}, see Equation 2) between the fixations in two scanpaths (A and B) varies with their length. Each data point is derived from 1000 randomly generated pairs of scanpaths. By chance, longer scanpaths with more fixations have a smaller separation and so are more similar.

Figure 4.1 illustrates two problems with the linear distance method. Firstly, the measurement does not take into account the temporal sequence of the scanpath (Figure 4.1C). Fixation locations are compared to whichever fixation is closest, regardless of when it occurred. As it ignores the order information, this metric would give high similarity for an example where in one scanpath the observer starts at the bottom left and works upwards whilst in the other they do the opposite. One way to avoid this problem might be to compute a "serial position" version, where the distance is computed between each fixation and that fixation which occurred in the same serial position in the other scanpath. However, this is skewed by any small deviations, as illustrated in Figure 3.1D. Differences between earlier fixations influence later comparisons by shifting the serial positions out of alignment.

The second problem occurs when the spatial distribution in one set of locations is very different from that in the other (Figure 3.1B). This leads to multiple fixations being compared to a single location in the other scanpath, potentially producing high similarity from two scanpaths that appear very different. Similarly, one outlier will skew two otherwise very similar scanpaths. Tatler et al. (2005) identify these problems, pointing out that the linear distance method is fundamentally confounded by differing amounts of spatial variability. Henderson et al. (2007) propose a "unique assignment" variant of the linear distance metric whereby each fixation is paired with just one other. All possible pairings are computed, and that chosen which minimises the average distance. A disadvantage of this approach, and of the serial position version, is that they require equal numbers of fixations in each set. As this would not be guaranteed in an experiment, a subjective decision would have to be made as to which fixations to include or reject. This variant can also be computationally complex due to the requirement to calculate distances for each permutation of pairings (two chains of n fixations can be combined in $n!$ ways, giving 40,320 permutations even when n is as low as 8).

4.2.2 String edit distance

To capture the temporal order of scanpaths, several researchers have utilised a method designed specifically for sequence analysis: the Levenshtein, or string edit distance (S. A. Brandt & Stark, 1997; Choi et al., 1995). This algorithm is an extension of the Hamming

distance that gives the difference between two strings of symbols in terms of how many positions are identical. The edit distance is defined as the number of operations (deletions, insertions and replacements) required to turn one string into another, and this distance decreases as strings become more similar. A method is available which computes the minimum number of operations required and this distance has been used for comparing a range of different items, from DNA sequences to birdsong (Sankhoff & Kruskal, 1983). The algorithm for computing the minimum distance is given in full in Brandt and Stark (1997). Figure 4.3 illustrates how the method can be applied to eye movement sequences made whilst viewing a natural scene. The visual stimulus is divided into regions, each of which is allocated a letter. Each 2-dimensional scanpath can then be transformed into a character string, and the edit distance between the two can be computed. It is often desirable to compare similarity across scanpaths of different lengths, so the distance is normalised by the number of fixations, and an index of similarity, which here I'll call *s*, calculated from its reciprocal

$$s = 1 - \frac{d}{\max\{n_1, n_2\}} \qquad (3)$$

where d is the edit distance between two scanpaths and n_1 and n_2 are the number of fixations within each sequence.

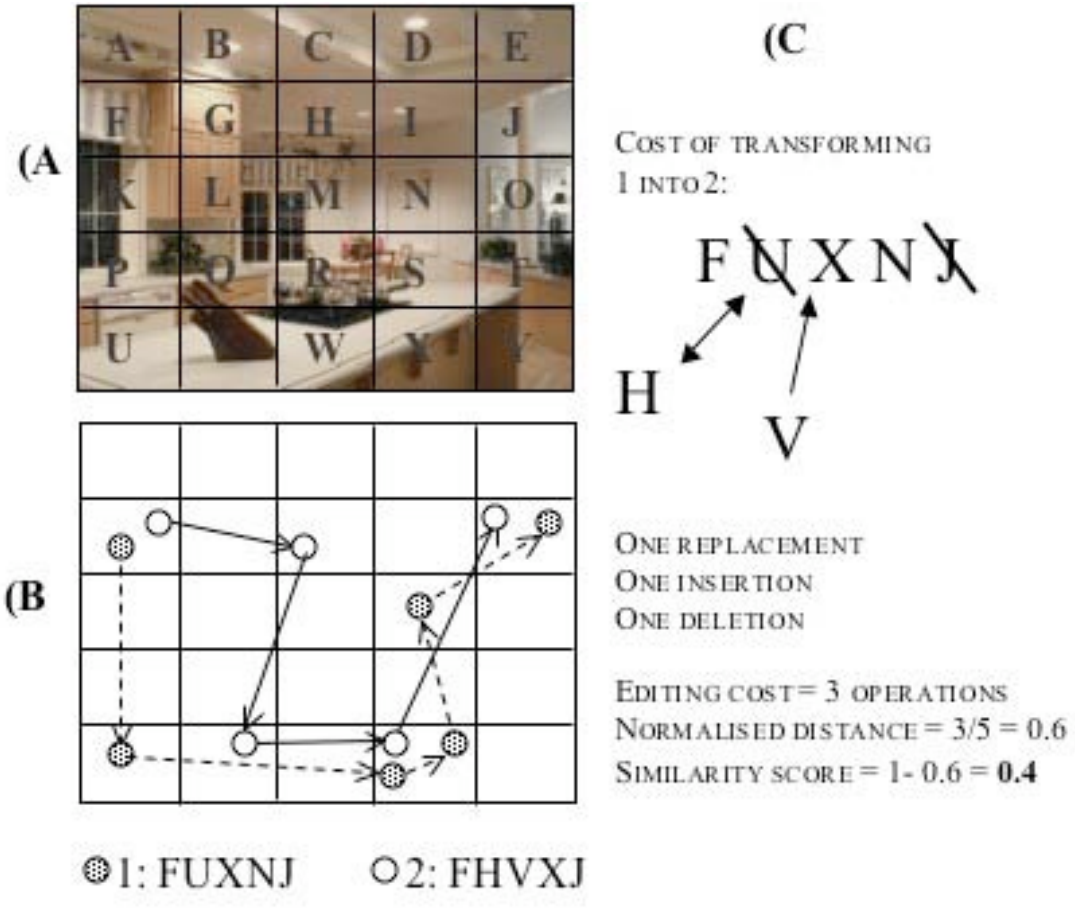


Figure 4.3. Using a string-editing procedure to evaluate scanpath similarity. A) The image is divided into regions, with each region allocated a character. In this case a 5 by 5 grid is used, though meaningful regions could be used. B) Scanpaths made whilst viewing the image are converted to a character string based on which region each fixation lands inside. Note that consecutive fixations on the "X" region (filled circles) have been condensed into a single character. C) The distance between two scanpaths is calculated as the minimum number of steps required to transform one string into another. This edit cost can be normalised and converted into a standardised similarity score.

There are several decisions that the researcher must make if using the edit distance method. Firstly, how are the regions chosen? In some cases there are areas of interest that can be predefined by the researcher. These might correspond to areas of a display or particular objects in a scene. In other cases it might be desirable to look at scanpaths over the whole image and to use regions of a constant size. In this situation the image can be divided into a grid, as in Figure 4.3, although this raises the question of how large these regions should be. A third possibility is to use the fixation data themselves to produce the regions, using statistical clustering techniques for example (Privitera & Stark, 2000). It might be useful to repeat the analysis with several different regions, perhaps of

varying sizes. Choi et al. (1995) argue that the estimates of similarity that they give are robust, whether using 10 or 15 regions.

A second decision that is commonly made is to condense consecutive fixations on the same region into a single character. This has the effect of removing repetitions in the string. Groner, Walder and Groner (1984) make a distinction between local and global scanning, with the former consisting of small readjustment saccades, of arguably less theoretical interest. Coarser scale movements between regions may be more useful. Of course, in combination with decisions regarding the size and shape of regions this will have a profound effect on the resulting edit distance.

How can researchers calculate the significance of s? This problem amounts to comparing experimentally derived similarity with a chance estimate. The chance similarity can be calculated as the probability that any two characters will be the same. Thus for a 5 by 5 grid there is a 1/25 chance that any two fixations will be in the same region. The similarity of randomly generated scanpaths could be used as the denominator in an estimate of absolute similarity analogous to Equation 2. Randomly generated similarity varies with the size and number of regions (see Figure 4.4). The random model might also need to be adjusted to take into account other biases present in the experimental sample. For example scanpaths might be constrained to start or finish in a particular place, or biased towards the centre. These biases will affect similarity estimates.

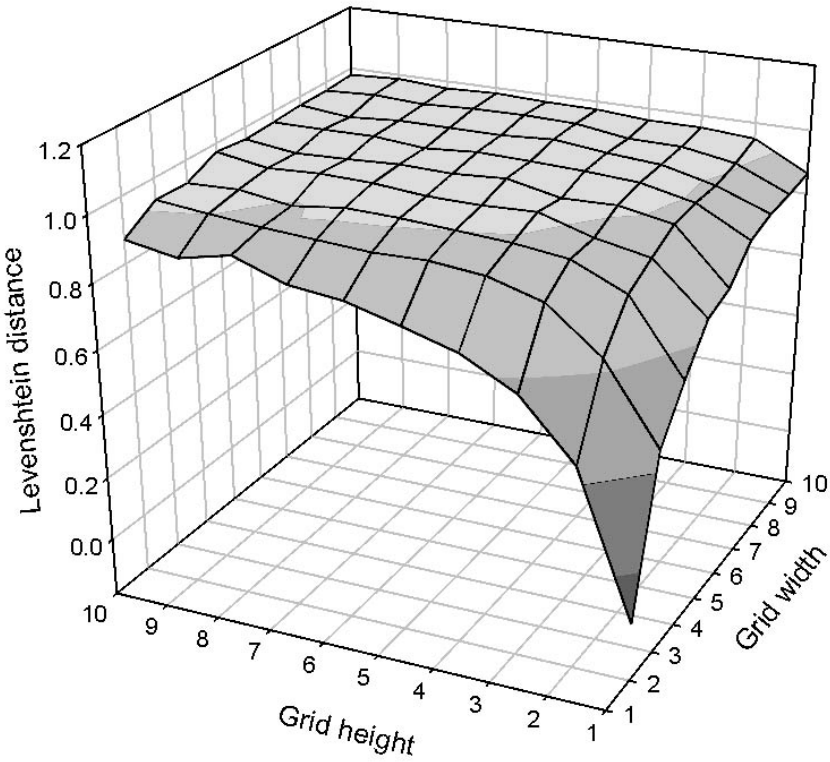


Figure 4.4. The normalised Levenshtein string-editing distance between the fixations in two randomly generated scanpaths varies with the size and number of regions. Data are based on scanpaths over an area divided into a grid of regions with dimensions varying from 1 x 1 (only one region where all fixations are evaluated as equal) to 10 x 10 (100 regions). Note that the distance expected by chance increases as a finer grid is used.

Although the string edit distance method successfully captures the temporal sequence of scanpath data, it reduces all spatial information to a binary choice where fixations are either in the same region or they are not. This is unsatisfactory and leads to some comparisons being equivalent despite large differences. The regions used, often a fairly arbitrary decision, are critical, as a fixation which lies just over the region border will be counted the same as one which is much further away. In an extreme case, a fixation might be computed as more similar to a location that lies in the same region than one that is closer but outside the region's bounds. In one sense, using more regions provides a more accurate representation with a higher resolution of the movements made. However, more regions also make the analysis less tolerant of small deviations that might otherwise seem negligible.

4.2.3 Combining linear and edit distance

Either linear distance metrics or the edit distance approach might be preferable in capturing the similarity required in any particular analysis. Privitera (2006) discusses three different measures: sequential similarity, locus similarity (giving an index of the positions both scanpaths dwell on, regardless of order and analogous to a linear distance method) and transition similarity (which consists of Markov matrices of the transitions between regions, see next section). An ideal method would combine metrics in a robust way. The edit distance can be adapted by weighting the operations involved, for example by penalising replacements more than insertions. In the standard algorithm all operations (replacements, deletions and insertions) incur an equal "cost" of one. In the case of scanpaths, this weighting could vary as a function of linear distance. The cost of replacing a character could be 1 if the fixations in question are very far apart and approach zero if they are very close. Similarly, although the cost of transforming between identical characters is normally zero, it could be greater if the two fixations, though within the same region, are further apart. More formally, the cost, c, of editing between two fixations i and j can be computed from their linear separation

$$c = \frac{d_{ij}}{\sqrt{(a^2 + b^2)}} \qquad (4)$$

where d_{ij} is the linear distance between i and j, and a and b are the dimensions of the display. Thus substituting a fixation for one on the other side of the display will give a greater cost than replacing it with one from the next region.

4.2.4 Additional methods

There are several other ways of analysing scanpaths. Some researchers have used Markov matrices that show the transition probabilities from one region to another (Stark & Ellis, 1981). While this may be useful for short scanpath segments, the matrices explode exponentially when longer chains are explored, making them impractical. Accordingly, in an analysis of scanpaths of drivers, Underwood et al (2003) used Markovian transition probabilities only for two-fixation and three-fixation sequences.

This method also requires decisions regarding which regions to use, and this risks the arbitrary division of visual space.

Fixation maps are a useful way of displaying eye movements, particularly those from large populations (Wooding, 2002; see Figure 4.5). In these maps, fixations are represented by a two-dimensional Gaussian centred on the fixation location. The width and height of the Gaussian can be varied, and multiple fixations summed, forming an "attentional landscape". Comparing two fixation maps is then possible. This may be an efficient way of computing the spatial similarity between two scan patterns which avoids some of the confounds associated with linear distance. Two maps could be correlated or a difference map could be produced (perhaps after normalising the height of the peaks). Spatially identical scanpaths would give a completely flat difference map. In theory, fixation maps hold no information regarding sequence, although it is possible to introduce a temporal element, either by combining maps derived from different time periods or by varying the height of fixation peaks over time. The fixation map approach also provides a way of identifying regions of interest from the data (for use in the edit distance method, for example).

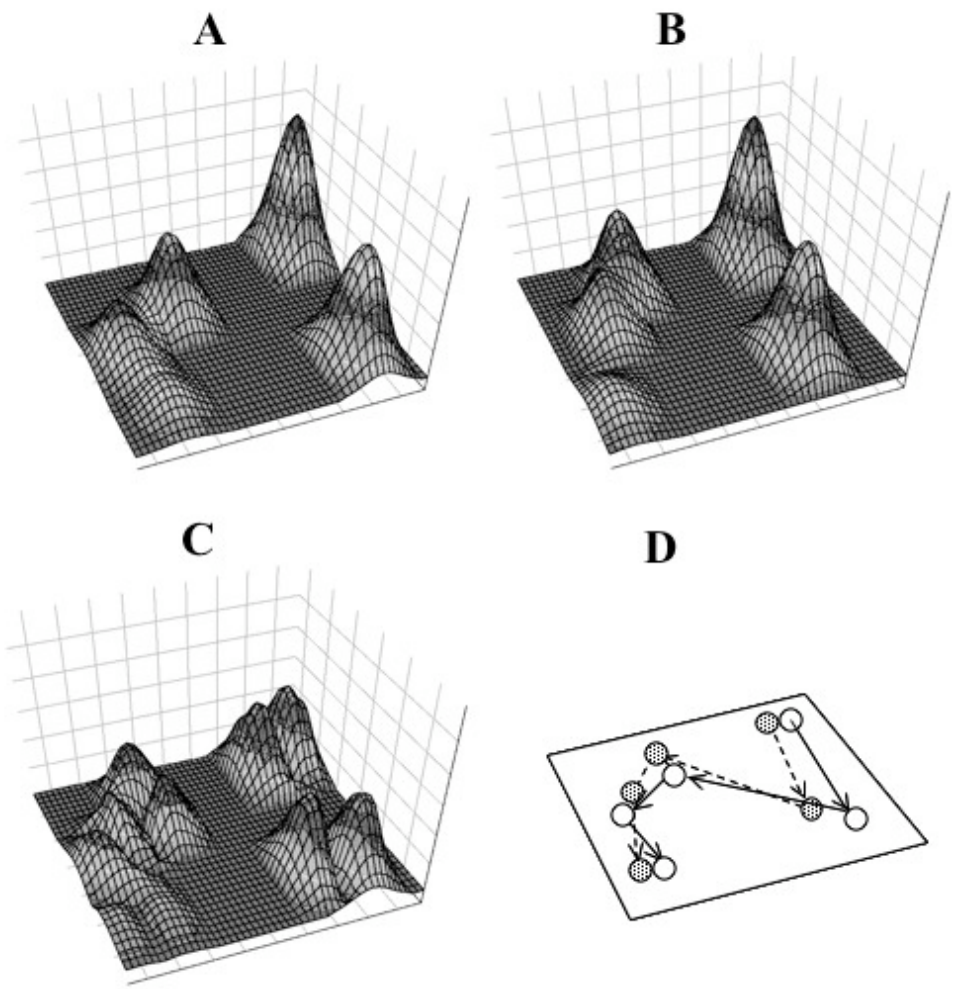


Figure 4.5. Fixation maps can be used to represent and compare scanpaths. Each fixation is represented by a 2D Gaussian. The z-axis (height) could represent fixation duration or another index. Alternatively, in order to encode sequence, height can represent when the fixation occurred. Early fixations produce a high peak, whilst later ones give a progressively lower distribution. Multiple fixations on the same area are summed together, producing an attentional landscape. A and B show two similar, hypothetical scanpaths, with z normalised to range from 0 to 1. The absolute difference between the two can be represented as a "difference" map (C). Identical scanpaths would show a totally flat difference map, whilst peaks indicate areas where fixation allocation differed between the two. A 2D schematic of the two scanpaths (A, open circles and B, filled circles) is also shown (D).

Tatler et al. (2005) also look at the full distribution of fixations across the image. In their method fixations are binned into 2˚ by 2˚ squares and a spatial probability distribution derived. The difference between two sets of fixations is then given by a measure from information theory, the Kullback-Leiber divergence, which computes the difference between the corresponding probability distributions. The Kullback-Leiber divergence gives the number of additional bits of information needed to describe one distribution given another. A low value indicates similar scan patterns. A disadvantage

of this technique is that it requires large amounts of data, so it is best used when the fixations from many trials and observers are being examined.

Finally, it should be noted that the measures considered in this chapter have considered only fixations, their location and temporal order. Researchers have often been interested in fixations, as this is where the processing of visual information occurs. Different similarity measures might be constructed which use instead the amplitude and angle of saccades as their raw data. The fixations could even be ignored altogether, making a trace which proceeds in a constant direction from A to C via B equivalent to one which goes straight from A to C. The careful analysis of many saccades during a task can also give information about the scanpath or the systematicity of eye movements. Gilchrist and Harvey (2006) draw conclusions about task strategy based on the angular distribution of saccades (for example the proportion of horizontal and vertical movements). I look at the distribution of saccade directions in natural scenes in Chapter 10. The current discussion has also assumed that all fixations are treated equally, and has ignored their duration. This information could be included in several of the measures, for example as a component in Equation 1 that gives more weight to longer fixations.

4.3 Conclusions and links forward

This chapter discussed two main approaches for comparing scanpaths. Distance-based methods are useful for summarising commonalities in *where* people look, though they are confounded by differences in spatial variability and do not reflect the temporal order of scanpaths (*when* people look there). The edit distance approach captures sequence at the expense of spatial resolution and requires decisions such as which regions to use. A combined algorithm that gives the weighted edit distance may capture similarity most efficiently. When is similarity reliable and worthy of further study? The scores generated by these methods should not be taken as absolute indications of scanpath similarity. Purely random estimates of chance tend to be too liberal and fail to take into account biases in the eye movement record. As a result similarity estimates are more useful when compared between experimental conditions to identify the relative significance of a scanpath comparison.

The similarity metrics can be used to answer questions such as “are two scanpaths made by the same person more similar than those made by two different individuals?” (that is are they truly idiosyncratic?) and “are scanpaths more similar when people view an image for the second time than when they view two novel displays?”. In terms of evaluating a saliency map model, the methods for comparing scanpaths give a different way of testing whether a fixation-selection algorithm which scans scene elements in order of saliency produces predictions which are similar to the scanpaths people actually produce. These methods will allow the sequential and spatial elements of scanpaths to be quantified so that they can be built into producing better models of natural vision.

The next chapter looks at the scanpaths produced by people viewing scenes in preparation for a memory test, and again when they are recognising them. Experiment 1 suggested that saliency did make a difference in where people looked in such a task. Chapter 5 looks in more depth at what saliency map models can predict about where people look, using analysis of spatial fixation locations and the methods of scanpath comparison discussed in this chapter.

5 Saliency and scanpaths at encoding and recognition

5.1 Introduction

In Experiment 1 I compared the attention to salient objects in different tasks, and found that in a "memory encoding" task, where there was no particular target, saliency made a difference. In this chapter I will look at the link between saliency and eye movements in a memory task in more depth. Several researchers have previously looked at the correlation between model-predicted saliency and where people look, but their estimates vary (Parkhurst et al., 2002; Peters et al., 2005; Henderson et al., 2007; see Chapter 1). Even when free-viewing static scenes (when top-down constraints are thought to be minimised) the correlation between saliency and fixation locations tends to be small; Parkhurst et al. (2002), for example, estimate mean correlations of 0.55 and 0.45 for fractals and natural photographs respectively. Here, I investigate the predictions of the saliency map model in terms of both the individual regions that are selected and the resulting scanpath, using some of the methods discussed in the previous chapter.

An important point when considering the results is that they are based on correlations. It is unclear to what degree fixations are actually caused by visual saliency. In a landscape photograph, for example, people may fixate the horizon due to saliency capturing their attention or due to top-down biases and 'gist' which include the assumption that informative objects such as people and buildings tend to be located near the horizon. One way to unravel these factors is to take a more experimental approach, as I do elsewhere in this thesis. In Chapter 3 the task being performed was important: saliency was only a significant factor in a memory task (where participants have to encode a scene for recognition later) and not in a search task. A memory-encoding task is useful as it encourages viewers to scan the scene naturally, paying attention to details, but without biasing them towards any particular feature. One motive for the current study, therefore, is to investigate in more detail the role of saliency in a memory task with natural scenes. By recording eye movements both at encoding, while the participant tries to remember the scene, and at recognition, I can look at the relationship between saliency and fixation with different task demands.

A particular issue with the correlational nature of some of the research mentioned above is that fixations are not distributed evenly throughout the scene but tend to be biased towards the centre of most displays. If saliency is also biased in the same way, model- and human-generated fixations may coincide but have no meaningful relationship. Tatler et al. (2005) show convincingly that the decrease in saliency over multiple fixations on a scene reported by Parkhurst et al. (2002) is an artefact of central biases in saliency, combined with a tendency to fixate in the centre that decreases over time. This demonstrates that it is essential to control for systematic biases in eye movements. In this chapter, I investigate such biases in more detail.

The Itti and Koch (2000) model treats all parts of the visual scene equally when computing saliency. The exception to this is that the previously selected location is inhibited, making it much less likely to be fixated again. This "inhibition of return" (IOR) also slightly enhances regions near to this location (due to the excitatory lobes on the IOR kernel), making small shifts more likely than large ones. Human eye movements may contain other biases, such as a left-to-right pattern of scanning, which might produce artefactual correlations. On the other hand, perhaps by building in relatively simple spatial or sequential biases, saliency map models might be able to account for much more of the variance in attentional allocation. For these reasons the current paper looks at both individual fixation locations and fixation sequences (scanpaths) and compares them to those generated by the model.

Chapter 4 discussed the comparison of visual scanpaths and its role in "scanpath theory" (Noton & Stark, 1971). This theory argues that eye movements are generated top-down, particularly in response to a previously seen image. Demonstrations that the scanpaths made by a viewer whilst encoding and recognizing the same image are similar have been used to argue that the scanpath made is encoded along with the visual features it explores (Stark & Ellis, 1981). Scanpath theory suggests that scanpaths are generated almost completely top-down, a product of the mental model of the observer. In contrast, the saliency map model argues that scanpaths can be explained in terms of bottom-up discontinuities and some simple selection processes which move between these salient points. The specifics of scanpath theory have rarely been replicated and this has led to a

decline in interest in studying fixation sequences (Henderson, 2003). One of the problems with scanpath theory was that it had difficulty accounting for the variability in scanpaths across viewings and observers. A strong form of scanpath theory would predict identical eye movements on the second exposure to an image, but it is interesting that a completely bottom-up saliency model would predict the same thing. If it is assumed that there is some variance between people and viewings (that is that neither model can fully account for eye movement sequences) then it becomes important to ask how much of the scanpaths people make are explained by previous viewings or saliency.

This chapter will investigate the relationship between scanpaths at encoding and recognition. It is an interesting question whether, if saliency models can explain a sequence of fixations on a scene, they might also be able to explain repeated sequences in response to the same image. In this case similar scanpaths are expected, not because they are encoded in any way, but because the image, and the saliency map, is constant. The saliency map model is sophisticated enough to predict actual scanpath sequences and my contribution in this chapter is to use these predictions to try and tease apart any effects of saliency or scanpath repetition.

It is also worth introducing some recent research by Althoff and Cohen (1999), who discuss the effect of memory or prior experience on eye movements. Their "eye-movement-based memory effect" was used as evidence of a global change in the sampling of visual information when scanning faces. This showed up as changes in the regions fixated when viewing famous versus non-famous faces and also in measures of the degree to which fixations were sequentially constrained. In particular, the scanpaths made when viewing famous faces were less constrained or systematic (as assessed by the 1st and 2nd order transitions between regions) than those made when viewing non-famous faces. Ryan, Althoff, Whitlow and Cohen (2000) subsequently investigated normal subjects and patients with amnesia viewing photographs of scenes and found decreased sampling of repeated scenes, independent of the patients' explicit awareness that they had seen them before. The emphasis here is on visual saliency, and hence the consequences which the saliency distribution might have on scanpaths at repeated viewings. The scanpath measures taken concentrate on similarity between viewings,

rather than a general assessment of constraint or systematicity. As memory is not the primary concern, memory performance and the time-course of recognition and any implicit effects on eye movements will not be discussed in detail. On the other hand, if prior exposure does produce a "reprocessing effect" then the sampling of visually salient regions should change. Althoff and Cohen (1999) suggest that the shift in processing found when viewing a face that is unfamiliar is designed to be optimally efficient at extracting information from a novel environment. As far as a saliency map represents where such information lies, then saliency might be more correlated with fixation in novel pictures than in previously seen pictures.

5.2 Experiment 2: scanpaths in a recognition task

In this experiment the eye movements of participants viewing photographs of natural scenes in preparation for a memory test are recorded. Whilst previously I have used this task merely to encourage scene exploration, here I investigate viewing of both previously seen and novel stimuli during a recognition test. Primarily, this study tests the hypothesis that the saliency model can predict fixation locations and sequences of these fixations, over and above any systematic biases. A number of other questions emerge. Does this relationship vary with the demands of the memory task (due to different top-down factors at encoding and recognition) or with novel and repeated stimuli at test (due to a repetition effect)? Alternatively, do the movements made at first viewing resemble eye movements on subsequent occasions and could a bottom-up model explain this?

5.2.1 Method

Participants

Twenty-one student volunteers with normal vision took part for payment. Inclusion in the study was contingent on reliable eye tracking calibration.

Stimuli, apparatus and design

A set of 90, high-resolution colour photographs of natural scenes was prepared as stimuli. Figure 5.1 shows some examples. All the stimuli showed exterior and interior scenes featuring houses, landscapes, furniture and other natural objects and were chosen to be similar to those within the same category. Of these pictures, half were designated "old"

and shown in both encoding and test phases, while the other half will be referred to as "new" and were shown only at test. Each subject saw the same set of old and new pictures, in a random order.

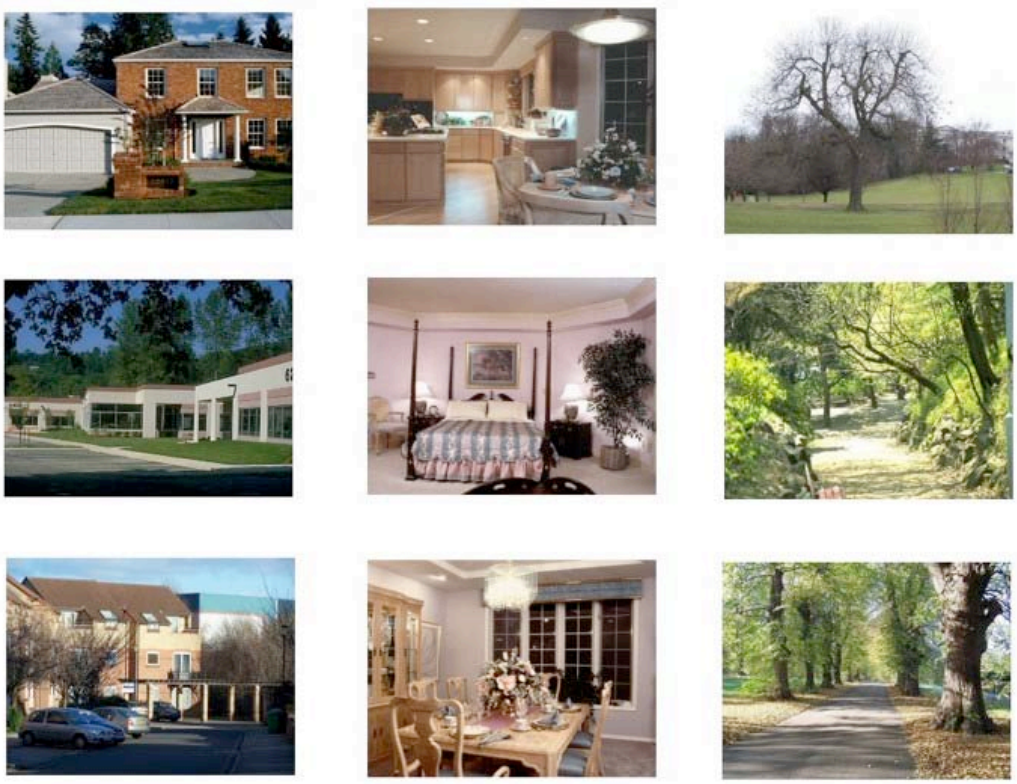


Figure 5.1. Some examples of the pictures used in the recognition experiment.

Saliency maps were produced for the first five simulated fixations and thus indicate the first five most salient regions for each picture (see Figure 5.2, top). The only further criterion for stimuli was that all 5 salient regions were non-contiguous; those pictures where the same or overlapping regions were re-selected within the first 5 fixations were replaced. In addition, a raw saliency map was computed for each picture to allow more comprehensive analyses of the saliency distribution present in the picture (see Figure 5.2, bottom, and results below for more details). These maps represent the combined conspicuity of the three feature channels, scaled to a fixed range of 0-255 and before control processes from the model (which favour a restricted number of saliency peaks and so promote a single "winner") are implemented.

The EyeLink I system was used with the standard set-up and calibration procedure.

Figure 5.2. Saliency map predictions for one stimulus from the experiment. The model produces a ranking or predicted scanpath (top) shown here as a series of circles linked by simulated shifts of attention. Also shown is a raw saliency map, produced by combining linear filtering at several spatial scales (bottom). Bright areas indicate regions of high saliency.

Procedure

Following calibration, participants were shown written instructions telling them to "inspect the following pictures in preparation for a memory test". In a practice phase designed to familiarise participants with the equipment, the displays, and the task, they were shown a set of six photographs with similar characteristics as the test set. They then viewed the same set, mixed with six novel photographs, in a random order. In each case they made a keyboard response to indicate "old" pictures they had seen before or "new", unseen pictures.

Following the practice phase, the experiment proper began. In the "encoding" phase all 45 encoding stimuli were presented in a randomized order. Each picture was

preceded by a central drift-correct marker and a fixation cross which ensured that fixation at picture onset was in the centre of the screen. Each picture was presented for 3 seconds and participants were free to scan the picture, after which the picture was offset and the next trial began.

After all 45 encoding stimuli had been presented the "test" phase began. During this phase all 90 stimuli (45 encoding stimuli which were now "old" and 45 not previously seen) were presented in a random order in exactly the same way as in the encoding phase and participants responded old or new using the keyboard. In order to facilitate an ideal comparison between encoding and test phases, each picture was again shown for 3 seconds and the response was recorded only if made during this period. This meant that although participants were encouraged to respond quickly they were able to inspect each picture for up to 3 seconds before responding, and the picture was not offset by their response.

5.2.2 Analysis and Results

How well did the saliency map model predict fixations and scanpaths in the different task phases? In order to answer this question I performed several different analyses. After describing the experimental data, I looked at the proportion of all fixations that were targeted at any of the first five "salient regions". Then, in order to look at model-predicted saliency values in more detail across the whole image and over multiple fixations I looked at the saliency value at fixation.

In both cases the results need to be compared against chance. One way in which to do this is to compare experimental data against a random model. For example if more human fixations than randomly generated fixations lay in salient regions then this would suggest the visual system is selecting regions based on saliency. However, a uniformly distributed random model might lead to a difference purely due to the systematic bias in eye movements toward the centre. For this reason a "biased random" model was also used where the random sampling of fixation locations is weighted by the spatial distribution found in the experimental data set. What other biases might affect the results? One way to look at the sequential aspects of the fixations made is to look at the 1st-order transitions between scene regions (see Descriptives and Figure 5.4). I therefore computed a third

model, a "random transition" model. This model sampled randomly based on the probability of moving to any one region from the current location. If participants tended to move in a clockwise direction, for example, this bias will be replicated in this model.

It is also desirable to look at full scanpaths and their sequence, rather than just individual fixations. Transition analysis is unsuited to this as the matrices involved explode exponentially when looking at sequences greater than two or three fixations. The previous chapter considered the background and methodology for comparing scanpaths, and here I chose two different methods: The distance-based algorithm developed by Mannan et al. (1995) was used to quantify the spatial distance between two scanpaths, whilst the 'string-edit distance' method looked at temporal sequence. Two main comparisons were made. First, were scanpaths at encoding similar to those made when looking at the same pictures at test? Second, were fixation sequences produced by the saliency model similar to the observed scanpaths? As with the other analyses a number of control comparisons were computed, which give an indication how much of any correlations are due to systematic biases.

Response data

In the test phase, participants responded to indicate whether the current stimulus was one they had seen previously. This task was performed well by all participants (percentage of correct trials, $M = 77\%$, $SD = 13\%$). Those trials that led to errors are excluded from all further analyses. Although it would be interesting to look at incorrect trials in further research (in the hope of relating fixation data to response accuracy) this was not appropriate here for two reasons. First, this comprised less than a quarter of the data and given the complexities of the analysis I was concerned about the statistical power available. Second, participants' behaviour in these cases is likely to be heterogeneous and could be interpreted as a true failure of recognition or as a result of confounding activity, boredom, fidgeting or any number of (potentially idiosyncratic) reasons.

The mean correct reaction time was 1459ms ($SD = 202$ms) and 1470ms ($SD = 212$ms) for old and new pictures respectively. This indicated that although the picture remained on the screen for 3s an average of approximately five or six fixations were

made before responding. As there may have been a change in scanning goals after responding, the eye movement measures looked principally at the first few fixations.

Fixation data

The raw eye movement data showed fixation locations and durations for each participant on each picture. In all cases, trials were excluded where the fixation at picture onset was not within 1° of the central region, which was ensured by presentation of the drift correct marker. The mean proportion of non-centred trials removed for this reason was 7.8% (*SD* =9). The first fixation was imposed by the experiment so was excluded from all further analysis. Unless otherwise stated, all statistical tests were repeated-measures ANOVAs performed on the participant means, with post hoc *t* tests (and Bonferonni correction) performed if necessary.

DESCRIPTIVES

The locations of all fixations (excluding the first which was necessarily in the centre) are plotted in Figure 5.3 and show a clear tendency for people to fixate in the centre of the display. This is not surprising, especially given that viewing started here, but reinforces the need to take into account this bias when evaluating model predictions. The first five most salient regions, across all stimuli, are also shown, and these are more distributed. In addition, to give an idea of any simple sequential patterns made by subjects, I calculated the 1st-order transition matrix for all fixations in the experiment. To do this the display was divided into a grid of 5 x 5 regions (this number achieved a balance between spatial resolution and ease of computation). The matrix shows the probability of moving from each of these 25 regions to any other and is depicted by a contour plot in Figure 5.4. The high probability along the leading diagonal indicates an increased likelihood to make small movements within the same region. People were also more likely to move from all regions into the centremost region, which is in agreement with the overall tendency to fixate there. Figures 5.3 and 5.4 were computed from large numbers of fixations (*N* = 27,252) collapsed across all stimuli, participants and task phases. These patterns make it clear that, across tasks and images, the visual scene is not sampled uniformly. In the next sub-section, three random models are explained which take account of this sampling.

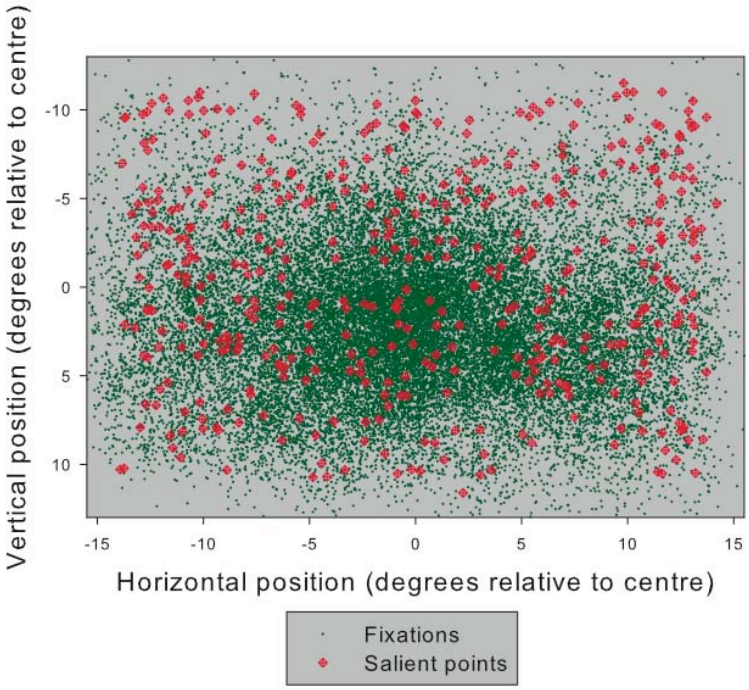


Figure 5.3. The locations of all fixations made by observers in the experiment, and the salient points, across all pictures. Fixations tended to be near the centre, while salient regions were distributed more evenly.

Table 5.1 summarises measures indicative of the general viewing behaviour at encoding and recognition. A repeated-measures ANOVA with three levels (encoding, old test items and new test items) showed no effect of task phase on the number of fixations, $F(1.2, 24.3) = 2.32$, $MSE = 0.36$, $p = .14$, Greenhouse-Geisser adjusted, although the effect on mean fixation duration approached significance, $F(1.3, 26.2) = 3.65$, $MSE = 632$, $p = .057$, Greenhouse-Geisser adjusted. New items at test elicited fixations that were around 15ms shorter on average than those on old items at encoding or at test.

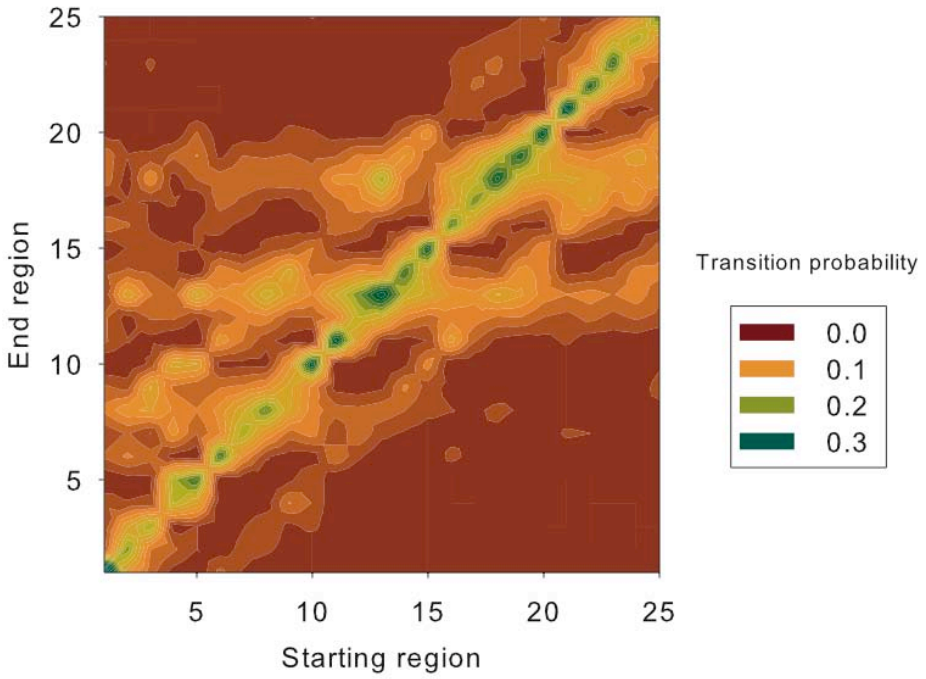


Figure 5.4. Transition probabilities across all fixations. The contour plot displays the transition matrix graphically, with each point representing the proportion of all fixations on the starting region (x axis) which moved to the end region (y axis). Note that fixations are most likely to move within a region, hence the high probabilities along the diagonal where x = y. Transitions from all regions are also more likely to move into the lower central regions, particularly regions 13 and 18.

	Number of fixations		**Mean fixation duration (ms)**	
	M	***SEM***	***M***	***SEM***
Encoding	10.59	0.31	278	12
Old items at test	10.47	0.27	279	11
New items at test	10.78	0.25	264	8

Table 5.1. Means and standard errors for general scanning behaviour in the different task phases.

RANDOM MODELS

In order to compare the predictions of the saliency model, three random data sets were generated containing the same number of fixations as the experimental data. The first set, "random", was produced by a uniformly distributed model which gave a random *x*,*y* coordinate for each fixation location. The second random model, "biased random",

weights the fixation location so that some locations are more likely than others. Specifically, the display was split into a grid of 25 sections and the probability of any one section being selected by the biased random model was equivalent to the proportion of experimental fixations in this region. Placement within a grid section was fully random. This model reproduces the central bias seen in Figure 5.3, and each modelled fixation is independent of the previous fixation. Henderson et al. (2007) also used a biased random model. A third model, "random transitions" duplicated the between-fixation regularities depicted in Figure 5.4. In this model the probability of a region being selected depended on the previously selected location (for the very first fixation this was the image centre). In each case the full range of possible regions was sampled based on the experimental probabilities of moving to each region from the previous one.

The use of these random data sets allows a more detailed test of the saliency map model. If human fixations are more likely to be targeted at salient regions than random fixations then this might be due either to general eye movement biases or to image specific bottom-up guidance. However, if this effect still remains when compared with the biased random and random transition models it cannot be attributed to general top-down biases, allowing a more confident endorsement of the saliency model. As mentioned previously, turning a correlational observation that fixations and saliency coincide into a causal statement that saliency causes fixation is problematic. Figure 5.3 shows that, across all images, fixations are not clearly tied to saliency but the remaining analyses will evaluate this more closely. The use of these random models to control for any across-image biases is a conservative effort to test for genuine saliency effects. However, being a correlational study the flip-side of this is that if these biases are actually caused by regularities in the distribution of saliency (because photographers place objects in the centre, for example) and are not just incidental, these analyses may underestimate the role of saliency.

PROPORTION OF FIXATIONS ON SALIENT REGIONS

As a first step to evaluating the saliency model, fixations were labelled as those on salient regions and those on non-salient regions. There were five non-contiguous salient regions on each picture, corresponding to the first five selected points in the Itti and Koch (2000)

model simulation (see Figure 5.2, top). These regions were defined as all points within 2° of the salient midpoint indicated by the model. The size of this region was chosen to reflect estimates of the size of the fovea and also the "focus of attention" parameter used in the model, which determines the region that receives IOR. A smaller region might give fewer salient fixations, but any noise arising from this decision will be replicated in the random models. For each trial, the number of salient fixations was compared to the total number of fixations made. The resulting proportion of salient fixations is shown in Figure 5.5 for experimental data as a function of task phase and for the random data sets. Across both encoding and recognition an average of 20% of all fixations landed on salient regions. The standard five salient regions comprised 10.4% of the area of each picture, and the random model produced salient fixations close to this rate.

There was a reliable effect of task phase, $F(2,40) = 75.6$, $MSE = 0.00015$, $p < .001$. A higher proportion of fixations were on salient regions in old trials than in either encoding ($t(20) = 8.71$) or new trials ($t(20) = 13.82$; both $p < .001$). Independent t tests then compared each task phase with each of the random models. In all cases the comparisons were highly reliable; For random, biased random and random transitions in order, at encoding ($t(40) = 10.8$, 7.1 and 8.1), for new pictures at test ($t(40) = 14.6$,9.5 and 10.8) and for old pictures at test ($t(40) = 16.8$, 13.2 and 14.2; all ps $< .001$). Participants made more fixations in salient regions than any of the random or biased models. This is clear evidence that the saliency model is better than chance at identifying regions which will be fixated.

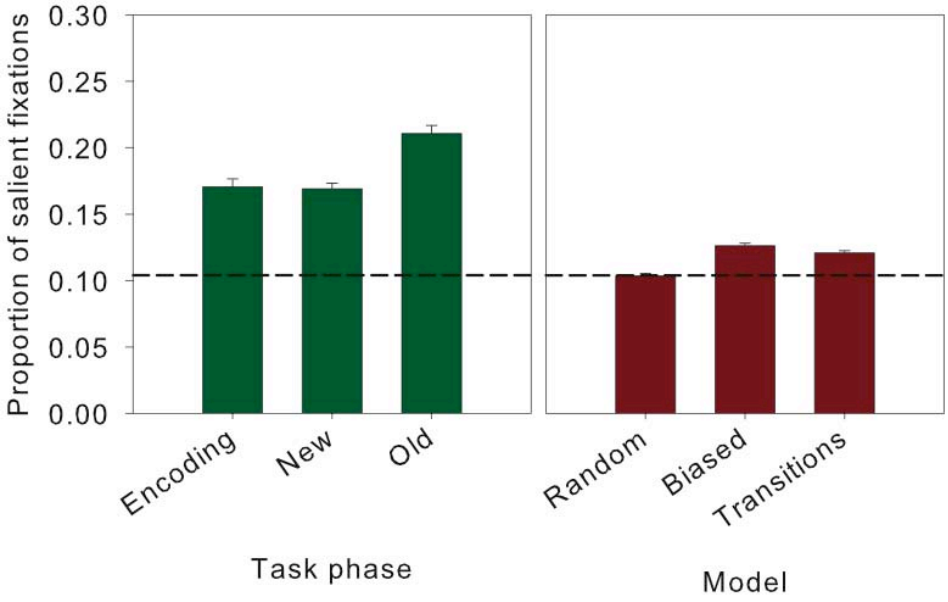


Figure 5.5. The mean proportion of all fixations that landed on salient regions, for observers in each task phase (left) and for the random models (right). Error bars show one standard error of the mean. The uniform chance expectancy (dashed line) is the value expected by chance if all locations were selected equally. This is equivalent to the proportion of the image covered by the salient regions.

SALIENCY AT FIXATION

The previous analysis looked only at the first five salient regions and this involved a relatively small subset of the fixations made. An alternative is to look at the saliency map of the whole image, which specifies a value for each part of the picture as opposed to just an ordinal rank. This allows the data from many more fixations to be analysed and so should lead to a fuller description of the relation between fixation and saliency. This analysis is also less dependent on the control processes of the model that simulate a focus of attention and IOR. Perhaps the underlying saliency map, rather than dynamic predictions of gaze shifts is more useful in predicting eye movements.

To begin with, values were extracted from a saliency map output from the model (see Figure 5.2, bottom). This map shows an intermediate stage in the saliency algorithm that highlights the combined conspicuity of all the features, but is produced before any inhibitory processes. A map from this stage was chosen so as to give a larger variance of saliency values across the image (the evolving saliency map at a later stage would be likely to reduce most regions to zero and enhance only a few peaks). The model scales all

the values to a fixed and arbitrary range between 0 and 255, and represents the map at 1/16 the size of the original image. It should be noted that there are arguments as to exactly how the saliency map should be normalised (see for example Parkurst et al, 2002). Here, these values were considered to be representative of the general saliency based model being tested.

Firstly, saliency values were extracted from fixation locations, scaled to 1/16, on the map for the corresponding stimulus. These were computed for the first five fixations following the first saccade after picture onset (note that this does not include the very first fixation which was in the centre) and averaged across trials. Five fixations was a reasonable number to include as at least this many would have occurred on all trials, and the reaction time data indicate that on average this many fixations were made before responding in the test phases.

As previously, the same analysis was carried out using the randomly generated data sets. If the visual system is indeed selecting locations based on saliency then the saliency values at fixated locations should be higher than at randomly generated locations. Alternatively, if people fixate regardless of saliency then the saliency values should not differ between experimental and simulated data. Figure 5.6 shows the mean saliency value at fixated locations. This value did not differ between encoding, new and old trials, $F(2,40) = 2.70$, $MSE = 13.7$, $p = .08$. Accordingly, the mean saliency value at the first five fixation locations was collapsed across task phases and the resulting grand mean was compared to the three random models. The participant grand mean (112.5, SD = 2.93) was reliably different from all three random models (independent *t* tests; $t(40) = 38.8$, 24.1 and 10.3 for random, biased random and random transitions respectively; all ps $< .001$). The results from this analysis, therefore, are entirely in agreement with those from the proportion of salient fixations; locations with high saliency received preferential fixation and biased random models did not fully capture this.

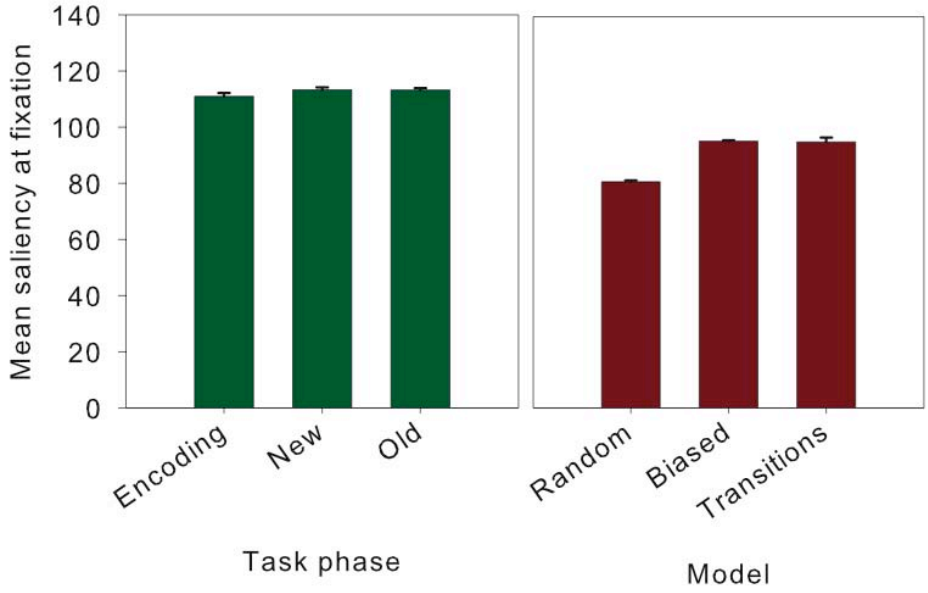


Figure 5.6. The mean saliency value at the first five fixation locations for experimental (left) and randomly generated (right) data. Error bars indicate one standard error of the mean.

A supplementary question is whether this advantage changes over the course of scene viewing. Parkhurst et al (2002) reported that the link between saliency and fixation decreased over the first few fixations. This was based on calculating a "chance-adjusted" saliency value that took into account the (unbiased) mean saliency value. However, if one assumes that fixations are biased towards the centre, high chance-adjusted saliency will occur without any meaningful relationship. Furthermore if the central bias of fixations decreases over time, as would seem logical if participants gradually widen their viewing, a drop in chance-adjusted saliency will occur artefactually (Tatler et al., 2005). Figure 5.7 shows the mean saliency value as a function of ordinal fixation number. A two-way, 3 x 5 repeated-measures ANOVA was computed for these data and as before found no reliable effect of the task phase $F(2,40) = 2.70$, $MSE = 68.5$, $p = .08$; saliency at fixation did not differ between inspection in the three task phases. There was a significant main effect of fixation number, $F(4,80) = 10.26$, $MSE = 41.9$, $p < .001$, which paired t tests revealed to be due to saliency at the second fixation which was significantly higher than the third ($t(20) = 4.44$), fourth ($t(20) = 5.05$) and fifth ($t(20) = 4.90$; Bonferroni corrected, all ps $< .005$, see Figure 5.7). There was a marginally significant difference between the

first and fifth fixations ($t(20) = 3.09$, $p = .043$) but no other comparisons were different and there was no interaction between task phase and fixation number, $F(8,160) < 1$. Thus although there is a slight suggestion of a downward trend this is almost entirely due to a high mean saliency value at the second fixation location.

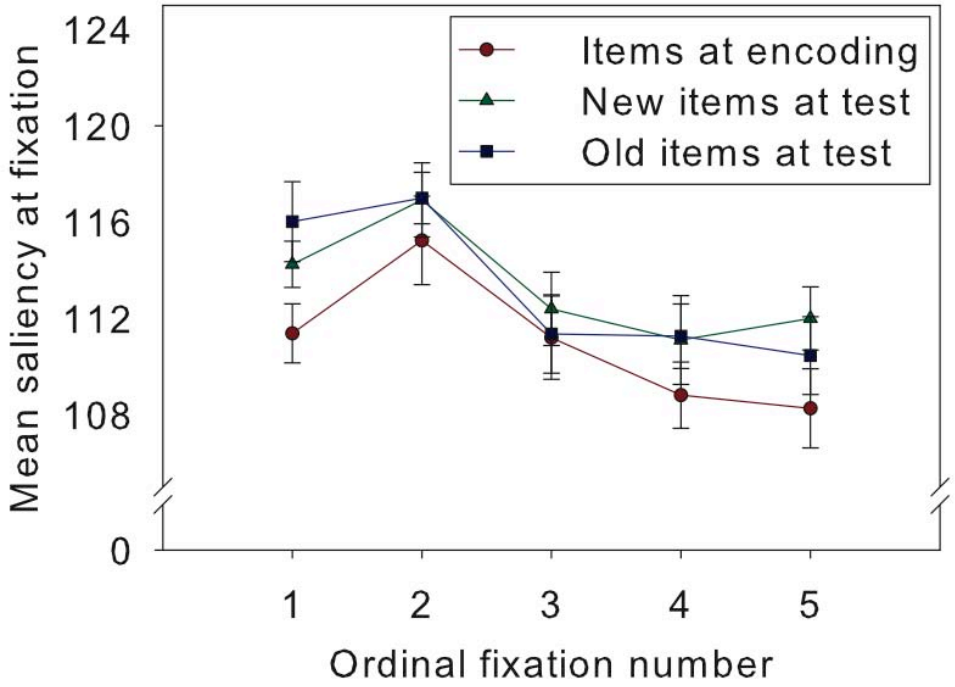


Figure 5.7. The mean saliency value in the three experimental phases, as a function of ordinal fixation number following the first saccade. Note that saliency remains high over several fixations and that the task phases show a similar pattern. Error bars show one standard error of the mean.

Scanpath data

The results so far have investigated the ability of the saliency map model to predict fixation locations, regardless of the order in which those locations are selected. An alternative way of looking at these locations is to consider the sequential scanpath that viewers make, and whether individuals produce repetitive scanpaths on multiple viewings of the same image (Choi et al., 1995; Noton & Stark, 1971). If systematic repetitions in eye movements are indeed a feature of natural scene viewing then characterising them could improve models, such as the saliency map model, which try to account for these movements.

In the next analysis, I first compared the sequence of fixations made by participants at encoding with the scanpath made when viewing the same stimulus at test. I then compared these scanpaths with predictions from the saliency model. The experimental data consisted of a variable-length sequence of fixations for each subject viewing each stimulus, giving a total of 135 x 21 = 2835 scanpaths (see Figures 5.8 and 5.9 for some examples). For the saliency map model, the first five predicted locations gave a "saliency scanpath" for each image. This is the scanpath that would be expected if, as the model suggests, the scene were scanned in order of declining saliency value.

Figure 5.8. Example scanpaths from a single subject when viewing a stimulus during encoding (a) and at test (c). A novel stimulus from the same category (a street scene) is also shown (b). In each case fixations are shown with circles (with diameter proportional to duration) and saccades with arrows. Scanning always started in the centre. (d) shows saliency output for the same stimulus, indicating a predicted scanpath.

This technique is described in detail in Chapter 4 and elsewhere (Brandt & Stark, 1997; Choi et al., 1995; Hacisalihzade, Allen, & Stark, 1992). To recap, it involves

turning a sequence of fixations into a string of characters by segregating the stimulus into labelled regions. The similarity between two strings is then computed by calculating the minimum number of editing steps required to turn one into the other. The resulting similarity index gives a value between 0 and 1 where 1 indicates identical strings. The images were divided into a 5 x 5 grid that was decided on after pilot analyses showed that it segregated the saliency scanpaths efficiently. The resulting 25 regions (rectangles of approximately 6.4˚ x 4.8˚) were labelled with the characters A to Y from left to right. Fixations were then labelled according to their spatial coordinates, resulting in a character string representing all the fixations made in this trial. The first, central, fixation was removed, and consecutive fixations on the same region (i.e. repeated characters in the string) were condensed into one. For more details on this method, and a discussion of some of the issues that arise from it, the reader is referred back to Figure 4.3 and the previous chapter.

How much similarity should be expected by chance? The probability that any two fixations will be on the same region is 1/25, although given the constraint that no two consecutive fixations will be on the same region the actual chance similarity will be slightly higher. A simulation run using a modified version of the Java program compared one thousand randomly generated pairs of 5 letter strings and gave an average similarity of 0.0417. Of course, the spatial biases previously discussed will also increase the similarity of scanpaths. For this reason a number of control comparisons were also computed which give a more useful baseline against which to measure similarity scores.

Although the string editing procedure quantifies sequence similarity, it does so at the expense of spatial resolution because the display must be arbitrarily divided into regions. To examine the overall spatial similarity of different scanpaths I used a method developed by Mannan et al (1995) for use with eye movements (see Figure 4.1 and Chapter 4 for this algorithm). This method computes the mean linear distance between a fixation in one scanpath and its nearest neighbour in the other set. Henderson et al (2007) refined this method slightly by adding the constraint that each fixation in a scanpath is assigned to only one other in the comparison scanpath. This “unique-assignment” (UA) version ensures that the scanpath comparison is not disproportionately affected by

differences in the overall distribution of fixations. This version therefore also requires that scanpaths have an equal number of fixations. Both the Mannan similarity metric and the UA version are standardised over the mean similarity of a randomly generated set of scanpaths with the same number of fixations as the tested set. This gives an index between 0 and 100 where 100 equals identical scanpaths and 0 equals scanpaths which are no more similar than chance. A negative index indicates scanpaths that systematically differ.

All three comparison metrics (string-edit, Mannan and UA) were used to analyse pilot data. The Mannan and UA metrics produced very similar patterns and as a result only the former is included in the remaining results.

COMPARING SCANPATHS AT ENCODING AND TEST

Figure 5.9 shows two example scanpaths from a participant looking at the same picture once at encoding and then again when recognising it at test. The similarity here is striking, and the comparison metrics discussed above aim to quantify this across all subjects and trials. To do this the scanpath for each subject viewing each image at encoding was compared to that from the same subject viewing the same (old) image at test. Figure 5.10 shows the mean similarity for both metrics.

Figure 5.9. Two scanpaths produced by one person viewing the same stimulus at encoding (left) and at test (right). Viewing started in the centre. Note that for the first five or six fixations the scanpaths are very similar.

There are several factors which might make two scanpaths from the same subject and stimulus on different occasions more similar than chance. If scanpaths are

idiosyncratic, similarity might be due to the sequence being generated by the same person with a general scanning strategy, rather than due to a sequence for that particular stimulus. In order to investigate this, each old test stimulus was arbitrarily paired with a new picture and the scanpaths on each compared. This comparison involves the same person performing the same task of recognition but with a different stimulus. Any similarity here therefore estimates the effect of a personal strategy, and cannot be caused by stimulus memory or constant bottom-up factors (as the stimulus in this case is different). In addition, pictures shown at encoding were arbitrarily paired with new pictures at test. This case involves the same person but viewing different stimuli under different task conditions (trying to encode the details of a picture and trying to recognise a previously seen picture). Similar scanpaths here would indicate a participant's similar scanning behaviour independent of stimulus and specific task instructions. Hence these control comparisons allow us to better interpret the similarity estimates found.

There was a significant effect of comparison source on string-edit similarity, (Figure 5.10a), $F(2,40) = 148$, $MSE = 0.00097$, $p < .001$. The similarity between encoding and old scanpaths was significantly higher than the similarity between old and new pictures ($t(20) = 13.14$) and encoding and new pictures ($t(20) = 15.27$, both ps $< .001$). The latter two comparisons did not differ, although all comparisons were greater than the randomly generated estimate of 0.0417 (one-sampled t test, all ts$(20) > 10$, all ps $< .001$).

An almost identical picture emerges with the Mannan metric (Figure 5.10b). The three comparisons were reliably different, $F(2,40) = 185$, $MSE = 18.6$, $p < .001$). Encoding and old scanpaths were more similar than the other two comparisons (old v. new, $t(20) = 11.65$; encoding v. new, $t(20) = 23.06$; both ps $< .001$). In addition, old and new scanpaths at test were more similar than encoding and new scanpaths ($t(20) = 4.49$, $p = .001$).

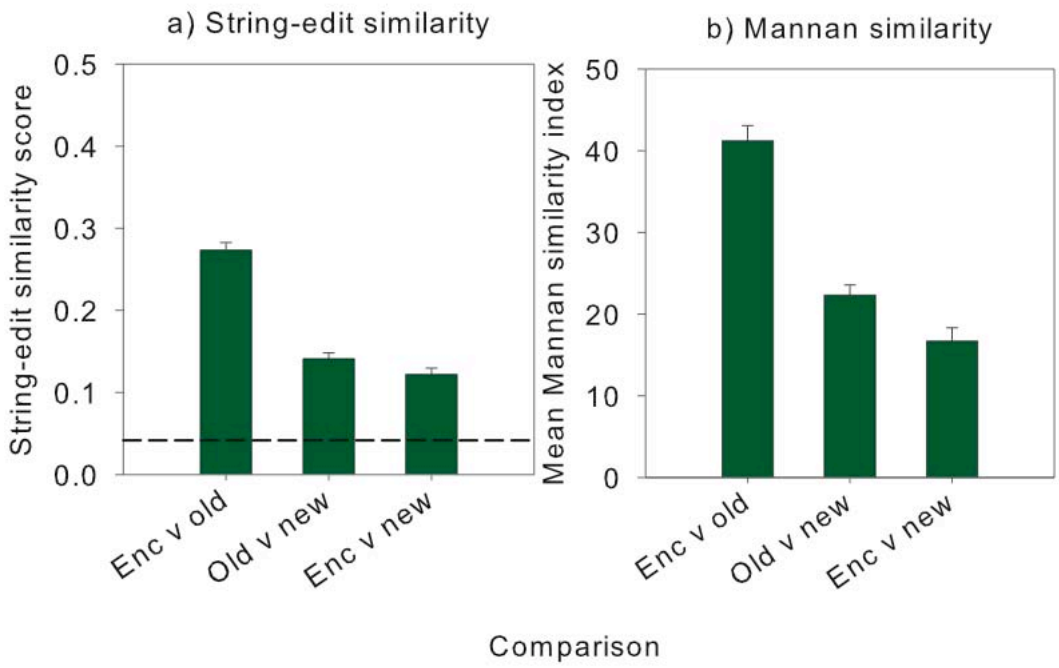


Figure 5.10. Mean similarity scores (with standard error bars) for the three comparisons and for each metric. The scanpaths in each comparison are from the encoding ("enc") phase or from "old" or "new" items in the test phase. The randomly estimated string-edit similarity is also shown (dashed line in a); the equivalent value in b) is zero.

COMPARING EXPERIMENTAL AND SALIENCY SCANPATHS

The scanpath comparison metrics provide a different way of evaluating the saliency map model. How well does this model account for the scanpaths generated by participants? The first five predicted fixations were taken as the "salient scanpath" and all experimental scanpaths were trimmed to the same length. The results are shown in Figure 5.11. It should be noted that, unlike the saliency analyses already performed, this method explicitly takes into account the sequence information provided by the model, rather than just the locations of independent fixations.

For the string-edit metric, there was no significant effect of task phase on similarity with the salient scanpath, $F < 1, p = .59$, indicating that any resemblance with the saliency model was constant at encoding and recognition. The similarity scores are noticeably lower than those in Figure 5.10 and are much closer to the chance estimate of 0.0417. The Mannan similarity scores are also lower for the saliency comparisons, although they are above zero, which indicates chance in this metric. This is not surprising as the analysis of fixations has previously shown that at least some of the fixated locations lie in salient regions and so should be spatially similar to the salient scanpaths.

There was also an effect of task phase on Mannan similarity with salient scanpaths, $F(2,40) = 15.9$, $MSE = 5.9$, $p < .001$. Scanpaths whilst viewing new items at test were more similar to saliency model scanpaths than those in the other task phases (encoding, $t(20) = 4.61$; old $t(20) = 4.49$, both $p = .001$).

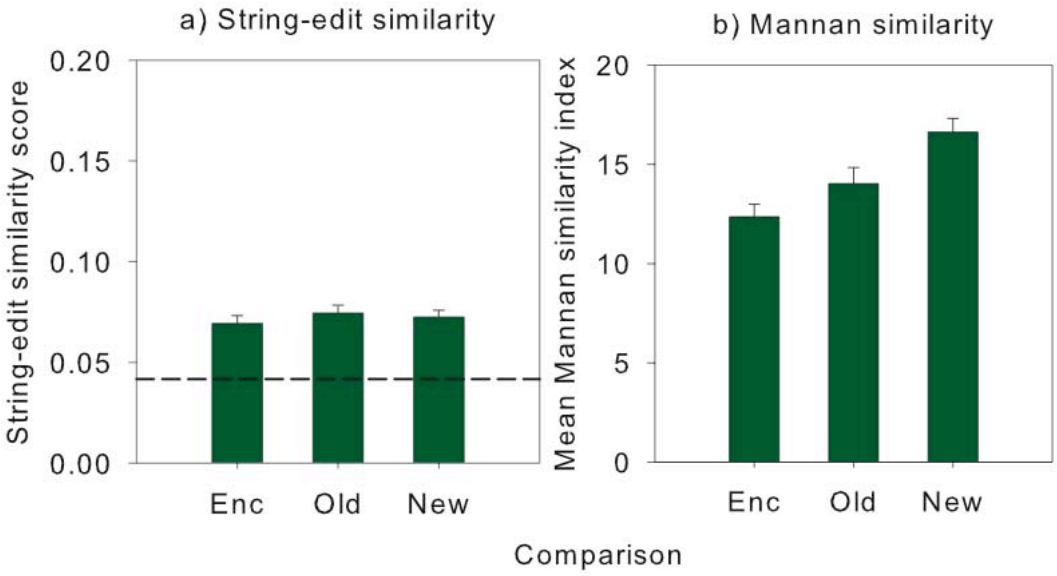


Figure 5.11. Mean similarity scores (with standard error bars) for the three task phases compared with the saliency scanpath and for each metric.

5.2.3 Discussion

This experiment compared model-generated and experimental eye movements, both in terms of the spatial location of individual fixations and their sequential order. It aimed to clarify how well the saliency map model can predict fixation locations and scanpaths. In addition it asked whether people repeat scanpaths on multiple viewings of the same stimulus. There were a number of interesting findings:

1. There was a tendency for fixations to target the most salient regions, as selected by the model. Biased models that took into account a central fixation distribution could not explain this result.

2. Across multiple fixations the saliency at fixated points was higher than predicted by chance and biased models. Saliency decreased slightly over multiple fixations but this was mostly due to elevated values on the second fixation.

3. This link between saliency and fixation did not vary significantly with the demands of the task (encoding and recognizing).

4. Scanpaths were most similar when compared between two viewings of the same picture by the same person (encoding v. old). This was reliable in terms of sequence, as evaluated by the string-edit distance, and linear distance. The similarity was significantly greater than control comparisons (encoding v. new and old v. new).

5. The saliency model-predicted scanpaths were not highly similar to human scanpaths

The results reported here are largely in agreement with those elsewhere that show that high saliency is predictive of fixation (Parkhurst et al., 2002; Itti, 2006; Underwood et al., 2006; Experiment 1). This relationship was evident in multiple findings in the present study: salient regions were fixated more often than chance and saliency values at fixation locations were consistently higher than average. The saliency map model was a reliable way of identifying areas likely to be fixated and it was much better than assuming a uniform distribution of attention. Importantly, it was also better than biased models that incorporated simple spatial and transitional biases. These biases explained more of the variance in where people look than purely random allocation, suggesting that these patterns need to be considered when looking at the relationship between saliency and fixation. Here, saliency had an effect over and above the central distribution of saliency and fixations. Other relatively simple biases, such as a predominance of horizontal saccades (see Chapter 10), might do better at explaining fixation locations and this is impetus for further research.

While this supports the general model, caution should be sounded regarding the correlational nature of these results. On their own, they cannot confirm that high saliency causes fixation or that the visual system is performing a saliency map based computation. Participants may have fixated regions based on other, top-down or bottom-up, factors which happen to coincide with saliency. When using natural images, where total control over all visual and task factors is impractical, both experimental manipulations (as in Experiment 1) and correlations are useful in testing a saliency map model. It should also be stressed that other, potentially simpler, bottom-up models that produce similar

predictions to the saliency map model might also outperform chance. When added to the biased random model simple information on the location of edges, for example, might perform as well as the saliency map model and render some of its other aspects unnecessary. I will consider this further in my conclusions in Chapter 12, but for the moment the relatively small difference between saliency model performance and chance is encouraging for those trying to explain this variance in other ways. For example, Tatler et al. (2005) find that edges and contrast are more related to fixations than luminance. Such research points towards variations in the feature channels that go into Itti and Koch's (2000) model. Aside from saliency maps, Raj et al. (2005) devised a model where fixation selection strives to minimize entropy or uncertainty in terms of local contrast. In theory this would be an efficient way of gaining spatial information in a memory task such as that presented here.

There are two further questions regarding the role of saliency in this experiment. Firstly, did the effect of saliency decline over multiple fixations? Although this is a popular framework for bringing together bottom-up and top-down factors in scene viewing, such that bottom-up saliency effects are gradually overridden by task knowledge, the findings here were not conclusive. The general shape of the saliency function over the first five fixations suggests a decrease, but even on the fifth fixation saliency was still higher than expected by chance. Although a significant change occurred this was wholly as a result of one high value on the second fixation. It is possible that the saliency map takes time to accumulate and therefore only becomes important at the second fixation, but analysis over longer durations would be needed to fully explore this effect. Tatler et al (2005) have demonstrated that decreases in saliency over time can be an artefact of central biases in saliency and fixation. Of course if salient regions are inhibited then shifts to less salient regions will become more frequent after the first few fixations. This experiment provides no clear evidence that there is any change in the bottom-up allocation of attention over scene viewing.

A second and related issue is whether saliency was predictive of fixation in all phases of the task. Henderson et al. (2007) argue that saliency cannot account for eye movements in a real-world search task involving counting the people in a scene, and

Underwood et al. (2006) and Experiment 1(a) found no experimental effect of saliency in search. In the current experiment, salient regions were fixated more often than chance and saliency values at fixation were higher than chance expectancy in all task phases. There was evidence that this was even more significant when old pictures were seen again during recognition.

There was a small effect of repetition on the mean fixation duration at test, such that people made longer fixations when viewing pictures they had seen before (with a fixed viewing time this implies fewer fixations as the two measures are inversely proportional, although the effect on fixation number was not reliable). These observations replicate findings by Ryan et al. (2000) and so support the idea of memory dependant changes in eye movements. Previous research has identified a reprocessing effect which leads to optimized scanning of novel items in comparison to previously seen items (Althoff & Cohen, 1999; Ryan et al., 2000). This might suggest that salient regions would be more potent at attracting attention in new pictures, as saliency might highlight more informative regions that it would be optimal to fixate. However, it was not the case that people looked at visually salient regions more often in new pictures than in old pictures at test.

The fact that saliency was predictive while preparing for a memory test is concordant with the memory condition in Experiment 1, and is reasonable for a task in which there were no pressing task demands as to what to look at. The demands when recognising pictures are quite different. It might be that participants are guided top-down to search out remembered features which could be defined, for example, by location within the picture (is there a car in the foreground? Is there a chimney above the roof?). A raw saliency map would not trigger these top-down shifts. However, there was no evidence that the recognition task was less driven by saliency. This might suggest that top-down shifts at recognition were not common, or that when they did occur they were made to remembered regions based on saliency. The relationship with saliency when viewing pictures for the second time suggests that memory independent of saliency was not a significant factor in eye guidance. Of course, if the first viewing is largely driven by saliency then the best remembered features will often be the most salient ones, making

the two factors difficult to unravel in this experiment. It has also been suggested that scanning, and the frequency of fixations in salient regions, could be explained by the semantic meaning of these regions. The interaction of saliency and scene semantics has been explored in some recent models (Torralba et al., 2006) and I will discuss it further in subsequent chapters.

It is worth noting that the saliency effects, while reliable, are not large. On average, only one of the first five fixations landed in a salient region, suggesting that there is plenty of variance in natural scene viewing unaccounted for. One aspect of oculomotor control that might explain some of this variance is temporal sequence information, and the present experiment explored this using the methods discussed in Chapter 4. Is the sequence of fixations made when inspecting a picture for the second time highly similar to that made on the first occasion? The present results suggest that scanpaths are indeed much more similar than would be expected from randomly generated sequences, and so support findings from simpler stimuli (Brandt & Stark, 1997; Noton & Stark, 1971).

Why is this the case? Scanpath theory (Noton & Stark, 1971) suggests that visual patterns are represented in memory as a network of features and the attention shifts between them. This network is then replayed and compared to the external stimulus when recognising the image later. By this account, the scanpaths at recognition were similar to those at test because they were stored and recalled top-down. The similarity seen here, though reliable, is much less than previously reported (Brandt and Stark report mean string-edit similarity as high as 0.75 in one subject), raising the question as to why there is still so much variance unaccounted for. Previous demonstrations of scanpath similarity have largely used simple patterns or line drawings, with fewer and larger regions of interest. It is likely that the much more complex photographs used here led to less scanpath repetition, possibly due to a greater influence of top-down scene knowledge. A closer analysis of the similarity seen here shows that a proportion (about half) of the resemblance can be explained by the consistent strategies and idiosyncrasy of individual subjects (as estimated by comparison with encoding and new picture similarity). These strategies need to be quantified further.

There remains significant similarity between first and second viewings, over and above this, that is to be explained. There are two different possible explanations. Similarity might be explained by a version of scanpath theory, or by some other top-down strategy. This strategy would need to be tied to the stimulus; otherwise it would also lead to increased similarity between old and new trials. It is possible that fixation locations were ordered in terms of decreasing semantic relevance, perhaps in combination with scene gist.

Alternatively, scanpath similarity could be explained by bottom-up allocation. By this account, scanpaths are similar because in both encoding and recognition phases fixation locations are at least partly determined by saliency which remains constant over viewings. In principle this seems likely as the analysis of fixation locations shows that participants were equally likely to target salient regions, which would remain the same, in encoding and old recognition items.

Top-down and bottom-up explanations produce the same prediction, partially similar scanpaths, and so are hard to distinguish from this alone. However, by using the saliency model and sequence analysis one can test explicitly whether scanpaths resemble the saliency model's predicted sequence at encoding and test. If this were the case then there would be no need to posit top-down scanpath generation. However, model-generated scanpaths were not highly similar to those shown by participants. It appears that the string editing method is particularly sensitive to sequence and that the model does not replicate the order in which salient regions are selected. This may explain the failure here to find a convincing change in saliency over multiple fixations. The similarity between scanpaths made at encoding and test was much greater than the comparison of either to the saliency model. Different models of bottom-up allocation might still explain repetitive scanpaths, but at the very least this finding suggests that such sequences are important for modelling eye movements.

5.3 Conclusions and links forward

This experiment offered partial support for a saliency map model. The model was much better than uniform or biased random models in marking out fixation locations. However, the model was poor at predicting the order in which these locations were selected. Interestingly, scanpaths were partially repeated on multiple exposures to the same stimulus, and my analyses suggest this is not due to saliency. Models may be missing out on sequential aspects of oculomotor control that could potentially predict fixation much better than saliency alone.

The findings from Experiment 2 therefore build on those from Chapter 3; fixation is associated with model-generated saliency, as other researchers have previously reported. This was evident both in terms of the regions that the model singles out as salient and which attract more fixations than expected by chance, and in terms of the underlying feature strength at each fixation location. It is important that extra steps were taken to show that these findings were not due to generic biases in fixation location and the distribution of saliency. Chapter 10 considers a different systematic bias in the saccades people make in an encoding task. When compared to biased random models the performance of the saliency model was somewhat less impressive than previously reported and there was no strong evidence for a decrease in saliency over time.

The scanpath comparisons put in to practice some of the methods discussed in the previous chapter. The results were interesting, showing that scanpaths might be useful for quantifying some of the top-down aspects of scene viewing and memory, and justifying the investigation of different methods for comparing them. The effects of saliency in a recognition task are also explored in Chapter 6, with the aim of relating recognition performance to saliency and attention at encoding.

6 Memorising and searching for salient and non-salient regions

6.1 Introduction

If saliency is important in attentional allocation then given the right task it should have an experimental effect on performance. In the real world, when people are stretched in terms of time or competing stimuli, what they fixate and pay attention to may become important, and predicting or modifying what this is based on visual saliency could be useful. Previous experiments in this thesis looked at attention to salient points in photographs in the context of either a memory task or a search task. The results found a reliable effect of saliency when viewers were encoding the picture for a later memory test but not necessarily when they were searching for a target. A possible criticism of Experiment 1, where objects were manipulated, is that placing objects in scenes might disrupt global scene properties or detract from the realism of the scene. In this experiment, unaltered photographs are used as stimuli and their regions are screened using the saliency map model. These regions can then be labelled in advance and probed with a memory task to explore attention towards them. Alternatively, these regions can be used as targets in a search task, allowing exactly the same stimuli to be used. This chapter describes two experiments using these tasks.

Previously a memory task, where participants are asked to view an image for a limited period "in preparation for a memory test", was mostly used to encourage unbiased scene viewing, and memory performance was not investigated. In Experiment 3, asking viewers whether a target region was present in a previously displayed picture tested their encoding of the scene. Two questions are considered. First, does saliency affect how soon and how often scene regions are inspected, consistent with previous findings? Second, does this effect later memory performance? In a task where cognitive resources are stretched saliency models would predict that people should spend more time inspecting regions with higher saliency and this may be reflected in better memory for these regions.

6.2 Experiment 3: encoding of scene regions in memory

6.2.1 Method

Participants

Eighteen student volunteers with normal or corrected-to-normal vision took part in return for payment.

Stimuli, design and apparatus

The stimuli for this experiment were 90 colour photographs. All the photographs were realistic scenes showing outdoor environments (landscapes, houses and other buildings) and interiors. These images were selected from a larger set following screening with the saliency map model. Some of the images had been previously used in Experiment 2. Figure 6.1 shows the first five locations selected by the model for one stimulus, along with the raw saliency map from which the model produces its output.

For the present experiment all of the stimuli had five non-contiguous regions as the first five locations (that is none of these regions were re-selected in the first 5 shifts of attention). In order to test attention to these locations, two regions were identified within each image: the most salient region ("S1") and the fifth most salient region ("S5"). This allowed a large number of potential stimuli to be generated without experimenter manipulation of objects or other interference. In each case the regions were squares of 200 by 200 pixels (approx 6˚ square) centred on the model-generated location. The size of patch was chosen following pilot observations showing that recognising smaller patches was much more difficult. These regions corresponded to areas that the model predicts should be potent at attracting fixations. If the model is able to predict dynamic eye movements across several fixations then S1 should be selected preferentially to S5. However, in some cases, this difference might be small (or even zero as the model will still decide between such regions due to small amounts of random noise which are added to the maps in the model). If the model is only useful in predicting the first few fixations, S5 regions may not be fixated preferentially at all.

To test the ability of the model to predict which regions will be inspected a further set of control regions was generated to give a baseline indication of the allocation

of attention to regions selected by chance. For each image a random (*x, y*) coordinate was produced and a region defined around it in the same way as with the salient regions. The final 90 stimuli were selected from a larger set with one further criterion: that none of the test regions (S1, S5 and Random control) overlapped, ensuring that fixation of each was mutually exclusive. An example of each type of region is included in Figure 6.1.

Figure 6.1. Stimulus screening and regions of interest. An example stimulus is shown, with predicted fixations from the saliency map model (top left). A raw conspicuity map for this stimulus, with lighter areas indicating regions of higher saliency is also shown (top right). The three types of probed regions, for this stimulus, are shown in the bottom of the figure.

As I have pointed out, studies may have produced artefactual correlations between saliency and fixation locations as a result of not accounting for central biases in fixation. If salient regions tend to be central then comparing them to fully random control regions might be misleading as higher fixation rates would be expected, not because salient regions are more potent at drawing attention, but because they coincide with a tendency to concentrate gaze in the centre of an image. It was therefore important in the current study that salient regions were not closer to the centre of the screen than control regions. In fact control regions were significantly *closer* to the centre (8.5˚ away on average) than S1 or S5 regions (10.22 and 10.37˚ respectively; paired-samples *t* tests, both *t*s>4, both *p*s < .001). The two groups of salient regions did not differ.

For the recognition test, regions were cropped and displayed in the centre of a blank screen with the instructions "was this region in the previous image?" Each stimulus was paired with a test region. In half of these cases the region came from the same image (target present trials), whilst the other half were filler trials where the region came from a different image in the same general category (for example both showed houses). In the 45 target present stimuli the three types of region (S1, S5 and control) were equally represented.

Pictures and regions were displayed and participants' eye movements were recorded during picture viewing using the standard EyeLink II set-up. Each participant was calibrated several times and validation indicated that the mean error was less than 0.5˚. Responses were entered via a gamepad.

Procedure

The procedure for each trial is illustrated in Figure 6.2. Following calibration, a short practice familiarised the participants with the task. Each trial began with a fixation marker in the centre of the screen, at which point the experimenter confirmed that fixation was maintained. The test image was then presented for 2 seconds whilst eye movements were recorded. This duration was used in a pilot experiment and adds some time pressure whilst allowing multiple shifts of attention. In this phase the participants had no reason to look at any particular region, other than preconceptions about what they would be tested on, and simply needed to encode the picture as best they could in the time available. Following this a test screen presented the probe region that may have been present in the previous image. Participants were asked to respond as quickly and accurately as possible whether the region was in the previous image by pressing keys marked "yes" or "no" on the gamepad. In order to encourage participants to memorize the images a further question asked where the region had been located.

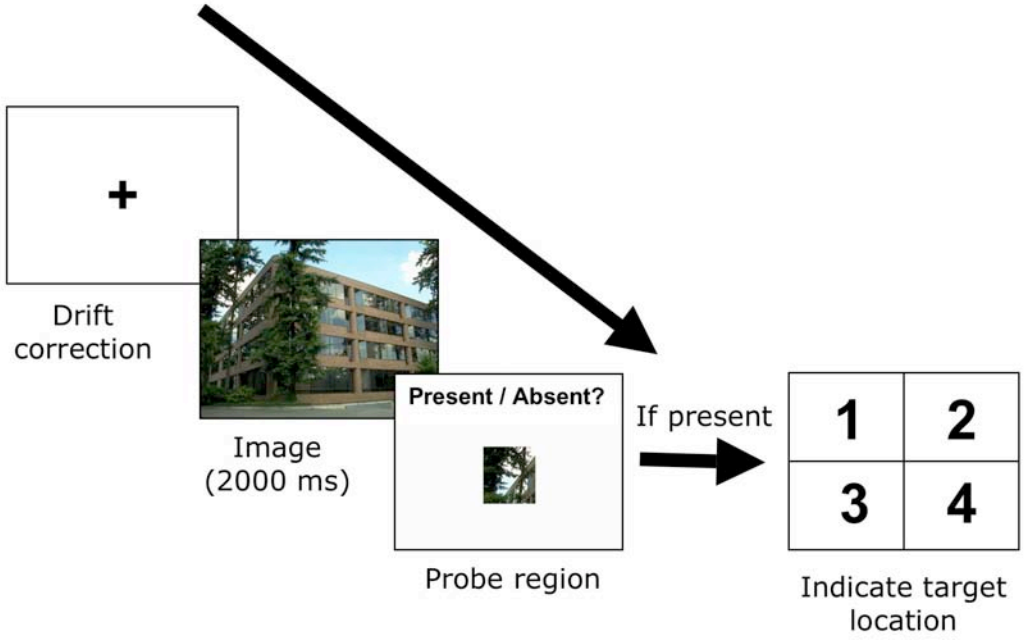


Figure 6.2. The procedure for one trial in Experiment 3. The task consisted of encoding an image and then recognising a small part of it.

Participants responded using one of four buttons indicating the four quadrants on the screen. This part of the procedure was added following a pilot study that showed that people often tried to base their response on the general gist of the scene rather than actually having seen the region. Thus the localisation task encouraged them to try and remember looking at that part. Some regions overlapped more than one quadrant, so participants were asked to respond with which quadrant most of the region had been in. The responses for this task were not analysed in detail. Trials proceeded for all 90 stimuli in a random order. I will continue to call this task a memory task, although given that visual recognition was tested immediately after encoding the image it should be noted that the system being probed may be better thought of as a sensory store than a long-term representation.

6.2.2 Results

The results are organised in terms of eye movement and manual response measures, in order to look for a difference in both the allocation of attention and subsequent memory performance. For the eye movement measures, the proportion of fixations in each region

gives an indication of the amount of processing of this region, whilst the first fixation time shows how early it was inspected. The measures taken are summarised in Table 6.1. In each case statistical analyses were repeated-measures ANOVA, with one within-subjects factor of region saliency consisting of three levels (S1,S5 or control). Pairwise comparisons, with Bonferroni correction, were performed post hoc.

	Region type					
	S1		**S5**		**Control**	
	M	***SEM***	***M***	***SEM***	***M***	***SEM***
Proportion of fixations on region *	0.43	*0.016*	0.422	*0.018*	0.306	*0.012*
Time to fixation (ms) *	810	*18*	796	*22*	871	*22*
Hit rate	0.852	*0.025*	0.822	*0.027*	0.87	*0.026*
***d* prime**	1.98	*0.19*	1.94	*0.2*	2.23	*0.19*
Reaction time (ms)	2187	*101*	2279	*82*	2252	*107*

Table 6.1. Means and standard errors for the measures taken in Experiment 3. Measures where there was a reliable effect of region type are marked with an asterisk.

Eye movement measures

Were salient regions inspected more often and earlier than non-salient, control regions during the encoding phase? There was a significant effect of region type on the proportion of fixations in a region, $F(2,34) = 31.78$, $MSE = 0.0027$, $p < .001$. Post hoc tests between the levels showed that control regions were fixated less often than either S1 or S5 regions ($t(17) = 7.7$ and 5.7 respectively, both ps $< .001$). S1 and S5 regions did not differ reliably.

There was also an effect of region saliency on the time to first fixation, $F(2,34) = 4.45$, $MSE = 6560$, $p < .05$. This is the time from picture onset until the first fixation on a region (trials where no fixation on this region occurred were not included). However, pairwise comparisons found that none of the levels differed reliably, although the

comparison between S5 regions and control regions approached significance ($t(17) = 2.6$, $p = .057$).

Response measures

Did the differences in attention toward different regions impact on the efficiency of the memory task? The first response measure, hit rate, looked at the proportion of all the target present trials answered correctly. There was no effect of the saliency of the probed region on hit rate, $F(2,34) = 1.47$, $MSE = 0.0072$, $p = .24$. In order to correct for possible biases and guessing in these responses, a measure from signal detection theory, *d* prime, was also computed. This measure takes into account the rate of false alarms (trials where the test region was not present in the original but the participant responded "yes"). *d* prime did not differ as a function of region saliency, $F(2,34) < 1$. Finally, the reaction time (to correct, target present trials) was inspected to see if participants responded quicker to salient regions than non-salient regions. Consistent with the other response measures, there was no significant difference in reaction time to regions with differing saliency, $F(2,34) = 1.22$, $MSE = 32729$, $p = .307$.

6.2.3 Discussion

This experiment looked at attention to predefined regions in natural scenes while encoding for a memory probe test, and also responses to this probe. The Itti and Koch (2000) model specifies that attention and eye movements should move first to the highest point on the saliency map. This model would predict, therefore, that S1 regions should be fixated earlier than S5 regions and control regions, particularly in a task where there is no particular impetus to look elsewhere. Presumably in some cases the less salient regions will not be fixated at all in the time available, and this might have translated into a difference in accuracy or reaction times.

The eye movement data revealed that while the most salient regions were fixated more often, they were not necessarily fixated earlier. There are several points worth noting here. The finding that a larger proportion of all fixations landed on salient regions than control regions is partial support for the model, suggesting that it does identify areas which are favoured in this encoding task. However, there was no reliable difference

between the first and the fifth most salient region in this measure. This means that while the model may be broadly useful in selecting regions which will be fixated more than chance (and so more than the control regions) the difference in ranking between the first predicted fixation and the fifth is negligible. In terms of time to target the first and fifth most salient regions did not differ significantly, again suggesting that the control processes of the model which move between peaks in the saliency map do not correctly predict the relative saliency of these regions. Indeed, both S1 and S5 regions were fixated after around 800 ms on average, or approximately three fixations, so the model predictions were not precise. The main effect of region saliency was mostly due to a marginally significant difference between the time to fixate S5 and control regions. The direction of the effect is consistent with the predictions of the model: that salient regions draw attention early.

The difference between the two eye movement measures is important. In a search task, Henderson et al. (2007) argue that higher fixation rates on salient areas might come about due to these regions being more likely to be refixated, due to semantic interest, rather than them capturing attention in the first place. A greater proportion of fixations on salient regions here might be explained by a tendency to see them as more informative or meriting more fixations as the contained more detail. Stronger evidence is the finding that such regions were fixated earlier and so presumably attracted attention in peripheral vision. However, this evidence was only marginally significant so the role of the saliency map in allocating early attention in this task is questionable.

Although salient regions were inspected more often, there was no effect of saliency on the accuracy or speed of participant responses. If one assumes that the probed region has been inspected on most occasions this is not particularly surprising, as all regions will be remembered equally regardless of when they were inspected. Indeed, regions fixated later on during encoding (and so closer to test), while presumably less "salient" in terms of attention, might even be remembered better due to a kind of recency effect. However, the probed region was actually fixated on less than half of the occasions it appeared. That people responded fairly accurately, even when not fixating the target region, suggests that the task was rather easy. In these cases the task may have been

possible due to information gleaned from the periphery during encoding, or based on memory for the global structure or gist of the previous scene. This makes further conclusions regarding responses to the memory probe problematic.

The findings of this experiment confirm previous data that model-predicted saliency can have an experimental effect in natural scenes, although this effect may not be conclusive in terms of the attentional mechanisms involved. One of the advantages of the stimuli and task used here is that it can be changed slightly to give a search task where the target is one of the regions described above. In Experiment 4 the same regions are probed but this time with a search task, in the hope of clarifying whether a target template or "cognitive override" operates regardless of target saliency.

6.3 Experiment 4: searching for scene regions

The attentional demands of a search task are quite different from those in Experiment 3. In an encoding task there is no strong expectation about what features will be present in a scene and what will be important to look at. In a search task the participant has some knowledge of what features will be important (those of the target) and perhaps some contextual information about where it may appear. What impact do these top-down factors have on the relationship between saliency and fixation? Unless the target is always the most salient thing in the display, control by an unweighted saliency map will hinder search because attention will move to salient items and will therefore take longer to move to the target.

Chapter 3 described an experiment that did not find a reliable effect of saliency on search. This is consistent with other studies that argue that visual search, particularly in natural stimuli, is dominated by top-down control. For example, Chen and Zelinsky (2006) found that irrelevant distractors were almost never fixated, even when visually salient. Presumably knowledge of target features provides a "cognitive override" of the low-level potency of conspicuous regions. This might be modelled by weighting target features in the saliency computation, and if this was totally efficient then target-dissimilar regions might receive a zero weighting. Pomplun (2006) extended experiments in simple visual search to look at top-down guidance by target features in complex greyscale photographs and patterns. This study demonstrated significant "saccadic selectivity"

across different feature dimensions (local intensity, contrast, spatial frequency and orientation). Participants were guided by the similarity of points in the image to the target patch in terms of these dimensions, and they also showed a "feature-ratio" effect (guidance by any one target feature was greater when it was less prevalent elsewhere in the display). This study did not measure saliency, but pointed out that purely bottom-up models of search are severely limited. However, if saliency is computed early and automatically then in the absence of other guidance salient targets should be easier to find than non-salient targets because they will have received extra "priming" by virtue of their low-level saliency.

There are some problems with extending these findings to natural stimuli. First, few researchers have been able to quantitatively measure saliency independent of target identity. In Experiment 1 the search task was performed very efficiently, with the target often found within a few fixations, and so generally low search times may have masked a difference between salient and less salient targets. Presumably, while top-down guidance (based on knowledge of target features) may dominate in a search task, highly salient targets will reflexively attract attention at least some of the time and so lead to faster search than non-salient targets. On the other hand, if saliency does not have an effect in a pressured search situation it is problematic for a model that considers that saliency drives early attention and suggests that effects of saliency in other tasks may not in fact be due to "capture" of attention.

6.3.1 Method

Participants

A group of sixteen students who had not taken part in Experiment 3 volunteered for this study and gave their informed consent.

Stimuli, apparatus and design

The stimuli and apparatus were exactly the same as those in Experiment 3. The search design broke down in the same way as that in the memory task. There were 90 trials, half of which were target-present trials. Three types of target region (S1, S5 and control) were represented equally.

Procedure

The procedure is depicted in Figure 6.3 and is very similar to that in Experiment 3, facilitating comparison between the two. Following a short calibration and practice, participants were presented with a series of 90 trials in a randomized order. Each trial began with a target region (which could be an S1, S5 or control region) presented in the centre of a blank screen for 2 seconds. This was followed by a drift-correct dot at which point the experimenter confirmed that fixation was steady and in the centre of the screen. A full size image then appeared on the screen and participants had to respond as quickly as possible to indicate whether the image contained the target region. The search display was terminated by their response, at which point a further screen asked them to indicate in which of the four quadrants the region was located. This test was added after a pilot study in order to encourage participants to search for and locate the target region specifically and not just respond based on overall gist or scene type.

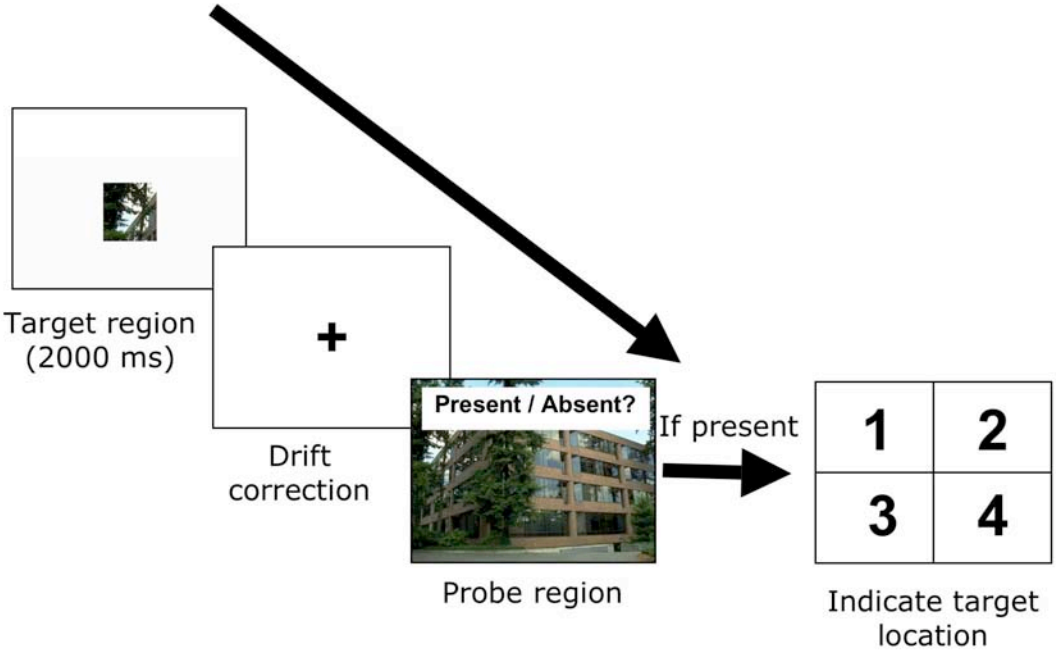


Figure 6.3. The procedure in Experiment 4 was a search task using the same probed regions as Experiment 3.

6.3.2 Results

As in Experiment 3 the measures of interest are both eye movement measures (how quickly and how often did participants move to a region?) and response measures (how quickly and accurately did participants respond that a region was present?). However, unlike the memory task, participants were cued with the target in advance and so the eye movement measures can be further split into cases where the region in question was the target and cases where it was not (because the target was a different region or the target was not present). All the measures discussed are summarised in Table 6.2 and they were analysed using repeated-measures ANOVA with one factor of region saliency with three levels.

	Region type					
	S1		**S5**		**Control**	
	M	***SEM***	***M***	***SEM***	***M***	***SEM***
Time to fixation (all trials, ms)	678	*39*	796	*80*	717	*52*
Time to fixation (when cued, ms)	685	*33*	776	*61*	745	*56*
Proportion of target absent trials where region is fixated *	0.414	*0.027*	0.405	*0.018*	0.248	*0.019*
Hit rate	0.892	*0.031*	0.846	*0.030*	0.879	*0.010*
***d* prime**	2.96	*0.22*	3.02	*0.23*	2.56	*0.23*
Reaction time (ms) *	2123	*165*	2633	*250*	2701	*215*

Table 6.2. Means and standard errors for the measures taken in Experiment 4. As previously, an asterisk indicates that there was a reliable difference between the regions.

Eye movement measures

How quickly did participants move their eyes to target regions of differing saliency? The time to target measure was computed for all trials and regions (regardless of which region was the target), and then separately for only those trials where the given region had been

cued in advance. In both cases only correct target present trials were included. Error rates were low (see response information below) and it was felt that behaviour in incorrect trials would be anomalous, so these trials were not included. There was no significant effect of saliency on time to target in all trials, $F(2,30) = 1.29$, $MSE = 44943$, $p = .291$, or only when cued, $F(2,30) < 1$. In both cases there were no reliable differences between time to target for S1, S5 and control regions.

An alternative way of looking at whether these regions attracted attention to different degrees is to look at those trials where the target region was not present. In these trials, which require a "no" response, the target template does not match any of the regions so can be only partly responsible for shifts of attention towards them. How often did regions receive fixation in these trials? The proportion of correct target absent trials where each type of region was fixated was entered into a repeated-measures ANOVA. There was a reliable effect of region saliency, $F(2,30) = 36.41$, $MSE = 0.0038$, $p < .001$. Both S1 and S5 regions were fixated more often than control regions in target absent trials ($t(15) = 6.5$ and 7.3, respectively, both ps $< .001$).

Response measures

The proportion of correct responses to target present trials ("hits"), *d*' and correct target present reaction times are included in the bottom half of Table 6.2 as a function of the saliency of the target region. Hit rate was generally high and did not vary significantly with target saliency, $F(2,30) = 1.10$, $MSE = 0.0082$, $p = .35$. *d* prime was also constant across conditions, $F(2,30) = 1.22$, $MSE = 0.84$, $p = .31$. However, there was a significant effect of saliency on reaction time, $F(1.3, 18.9) = 10.6$, $MSE = 237387$, $p < .005$, using Greenhouse-Geisser correction. Participants reacted quicker when searching for S1 regions than either S5 regions ($t(15) = 2.9$, $p < .05$) or control regions ($t(15) = 4.0$, $p < .005$). Response times to S5 regions and control regions did not differ reliably.

6.3.3 Discussion

The effects of saliency in this chapter were mixed. When the target region was present, salient regions were not fixated significantly earlier than control regions. However, overall reaction time was higher for S1 trials than either S5 or control trials. This

indicates that the model did predict a difference in the right direction in terms of responses, even in a natural search task. In addition, participants were more likely to fixate salient regions than control regions in target absent trials.

The findings of this study suggest that, given the right conditions, saliency can make a difference in a search task, which is contrary to some of the conclusions of Chapter 3. Henderson et al.'s (2007) "flat landscape hypothesis" suggests that search is not affected by saliency but is based purely on top-down guidance. The findings here count against this and indicate that saliency is not completely overridden by top-down target knowledge. Target features may guide the eyes within a saliency map framework. Such guidance has been shown to be efficient in naturalistic stimuli (Rao et al., 2002; Pomplun, 2006). Presumably if this top-down weighting modifies a saliency landscape then salient and control regions will both receive a top-down boost, but salient regions will still be represented by a higher "peak". If there are other saliency peaks competing for attention then at least some of the time these might be higher than a non-salient target region, particularly if they are also similar to the target, and so it will take longer for the visual system to select the target region and to respond. The failure to find a reliable effect of saliency on time to fixate the target, however, does not fully support this account, although there was a difference in the predicted direction. Interestingly there was a large difference in the rate of fixation in target absent trials, indicating that in the absence of accurate target guidance salient regions did capture attention more than control regions.

A number of questions emerge from these data. Why did saliency affect reaction time here when it did not when searching for objects in Experiment 1(a)? There are several important differences between the two search tasks that might explain this. Responses took longer, indicating that the task was harder, which may have allowed a difference to show. Rather than searching for whole objects, the search task here required searching for regions, which could be parts of objects or background features. It is likely that people are well practised at finding objects, and top-down expectations about where they are likely to occur might override saliency much more than in the region search task used here.

Why did participants react faster to salient regions when they did not necessarily fixate them earlier? This is problematic for the logic that saliency attracts attention and suggests that the locus of the effect on reaction time could be the speed of processing of the region rather than the time to find it. This might be because salient regions differed from control regions in terms of the information they contained, perhaps with salient regions being processed easier (and fixated for less time). It is striking that there is a large difference between the time to target measure and the manual reaction time. As the time to target measure looks at the very first fixation, this suggests that on most trials several more fixations were made before responding. If salient regions were more likely to be identified in the first few fixations, and not just skipped over, then this could underlie the reaction time difference. By this explanation control regions were fixated just as quickly but they required more refixations or were less likely to be identified first time than salient regions. One problem with the procedure, however, is that as participants were later asked to say *where* the region had occurred, lengthened reaction times to the initial detection task might actually reflect participants trying to encode region location. The two responses are hard to separate in this task.

6.4 Conclusions and links forward

Experiment 3 replicated the finding from Experiment 2 that region saliency had an effect on eye movements in a memory task; the most salient region was inspected more often, and earlier than control regions. Interestingly, in this task this did not have an impact on recognition, so that although salient regions did attract attention they were not necessarily encoded or recognised better. Changing the paradigm slightly gave a search task, and here saliency had clear effects on performance. The search target, a scene region, was different from the identifiable objects used in Experiment 1 and this may explain the discrepancy in the effect of saliency. Effects of saliency in search will be examined in a different paradigm in Chapters 8 and 9. First, the current search task will be extended in order to probe more deeply the cause of the advantage for salient regions. Is the saliency map truly identifying regions that are preferentially attended to, and which features cause this? The following chapter addresses this question.

7 Manipulating the information available in a region search task

7.1 Introduction

The previous experiment demonstrated that effects of saliency could be found in a region search task. However, they are somewhat inconclusive regarding whether saliency actually affects attentional orienting. In Experiment 3 there was evidence that salient regions were fixated earlier than control regions but this difference was not reliable in Experiment 4. The Itti and Koch (2000) model makes clear predictions that salient regions should be identified and represented (in a saliency map) with peripheral vision and so they should effect targeted eye movements and not just the processing at fixation. The model also specifies the features that are important in peripheral vision as contrast in terms of intensity, orientation and colour. Experiment 5(a) repeats the search task with gaze-contingent filtering in the periphery. Two further experiments (5(b) and 5(c)) look at search in inverted scenes and search with a fully masked display respectively. In each case the aim is to identify the information that is necessary for a saliency effect in search to emerge. Experiments 5(a) and 5(c) utilise a gaze-contingent display—one that changes in response to eye movement events—so I will begin by outlining how they have been used in research into scene perception.

Gaze-contingent displays have been used in a variety of different eye-movement studies to manipulate the information available at different eccentricities. In reading, the gaze-contingent, "moving window" design masks text outside a window that is centred on fixation. By varying the size of the window, the perceptual span at which reading can still proceed normally can be assessed (McConkie & Rayner, 1975; see Rayner, 1998, for a review). In visual search, a gaze contingent window has been used to manipulate or mask the target-similar features which are present in the periphery (Pomplun, Reingold, & Shen, 2001). Masking reduced the amount of saccadic selectivity, supporting the idea that visual guidance operates preattentively and in parallel across the display. In scene perception, multi-resolutional displays take this further by allowing the resolution of the scene to be steadily degraded as a function of increasing eccentricity. This has applications in terms of reducing the resources required to display a high resolution image

(by only displaying a small window at fixation at the highest resolution (Reingold, Loschky, McConkie, & Stampe, 2003). When the function relating eccentricity and resolution is lower than or matches that in the human visual system this is not detected by the observer (Loschky, McConkie, Yang, & Miller, 2005). The implication of these studies for eye movement control is that visual resolution limits the information that can be gleaned from the periphery and go on to affect saccade targeting. When above-threshold filtering is used, this should alter the way eye movements are made. An alternative to the moving window design, a "moving mask" can be implemented, which moves with, and disrupts the information at, fixation. How do the two types of masking effect eye movements? In terms of the duration of fixations made, Greene (2006) showed that peripheral masking lengthened fixations which were knowledge-driven. Van Diepen and d'Ydewalle (2003) found that while foveal masking affected fixation durations more severely (as would be expected if fixations are dominated by processing at the current location) peripheral masking also had an effect. Interestingly the effect of peripheral distortion on the current fixation had an effect even when this occurred only at the beginning of the fixation, suggesting that even early in fixation peripheral information is important, presumably for planning the next saccade.

The above results predict that gaze-contingent distortion will affect visual search and scene perception in terms of a general disruption leading to longer search times, shorter saccades and longer fixations. Reducing the visual information in this way is likely to hinder search performance, as people will not be able to move towards the target as quickly and efficiently as without it. The gaze contingent design allows visual information outside of fixation to be altered, whilst leaving that available at fixation intact. What effect will this have on the advantage for salient objects seen in previous experiments? If saliency affects shifts of attention then we should find clear effects on the time to fixate the target, and the gaze-contingent filtering may moderate this.

What does the saliency map model predict in a filtered display? The saliency of items in the periphery will be reduced relative to those things in the window of fixation. Loschky & McConkie (2002) suggest that peripheral filtering lowers the probability that an object will win the competition for saliency. In addition, the boundary of the window

might create a salient contour. Assuming that saccades within the window are suppressed, the saliency of the regions of interest will be affected differently depending on the type of filtering. For example, if colour information is important in computing the saliency map then in a greyscale display "salient" regions may no longer stand out. If such filtering removes the effect of saliency seen in Experiment 4 then it would suggest that colour is important in marking out salient regions and would reinforce the model as a predictor of attentional selection in peripheral vision.

7.2 Experiment 5(a): searching for regions in a gaze contingent display

7.2.1 Method

Participants

Three separate groups of volunteers were recruited, none of who had taken part in Experiments 3 or 4. The groups experienced different filtering conditions (blur, greyscale and contrast) and had 17, 18 and 17 people respectively. All participants were students with normal vision and several were replaced in order to assure good calibration with the eye tracker.

Stimuli, design and apparatus

Visual filtering was varied between groups, with each group receiving a different type of filtering. The same set of 90 stimuli was used, as described in the previous chapter, with target region manipulated within-subjects. Target regions were not modified. The gaze-contingent procedure used a small, 200 pixel (approximately 6˚) square window of the unfiltered pictures, cropped around fixation as the foreground to the search display. Although other studies have used circular or oval windows, in practice a square window is much easier to produce and leads to very similar results considering the current study was not concerned with small differences in perceptual span or the drop off with eccentricity. The background, which was shown outside of fixation, was taken from the same set of images, but was filtered in one of three ways (see Figure 7.1). Greyscale filtering consisted of changing the images from full colour to 256 levels of grey. The blur manipulation applied a Gaussian blur function (with a sigma value of 1˚) to the image,

giving a low pass filtered version that lost much of the fine detail. The level of blur was chosen to be relatively severe in order to make any effects clear. Finally, half contrast images were adjusted in order to the lower the global contrast by 50%. All manipulations were produced in a batch using the commercially available software Adobe Photoshop CS. The three modifications (greyscale, blur and half-contrast) were designed in order to reduce the information in the colour, orientation and intensity channels respectively in the Itti and Koch saliency map model.

The EyeLink II system was used again and automated the gaze-contingent display. The eye tracker recorded eye position at 500 Hz, and a CRT monitor presented images with a high refresh rate of 125 Hz. Both these values are within a suitable range to allow real-time updating of the display with limited lag between the eyes and the picture.

Figure 7.1. An example of one of the scenes used in the experiment (top) and the filtered versions which acted as the peripheral background in the gaze-contingent display (bottom). Three types of filtering were used: greyscale (left), half-contrast (middle) and blur (right).

Procedure

Participants were pseudo-randomly assigned to one of the filtering conditions, and the search procedure was very similar to that in Experiment 4. It is particularly important to achieve a good calibration with a gaze-contingent display, in order to avoid a mismatch between where participants are fixating and the moving window. As a result, participants were calibrated several times before the experiment began, and after each block of 30 trials. An extra practice session was given before the experimental trials in order to

familiarise participants with the display. Instructions informed participants that the display would be modified in response to their eyes, but they were told to ignore this and perform as well as possible. The experimental procedure then began with all trials shown in a randomized order.

In each trial one of three types of target region was presented, for 2 seconds, followed by a drift correct dot and then a full size image. The search display was presented using a "moving window" technique (see Figure 7.2 for a diagram). Fixation started in the centre of the display and an unfiltered window moved with the eyes, such that every time eye position changed, the display updated to centre the window on the new fixation location. The results are a smooth display where everything except fixation is distorted. As previously participants made a yes/no response on a keypad to indicate if the target region was present, and if their response had been yes they subsequently pressed one of four keys to indicate in which quadrant it had appeared.

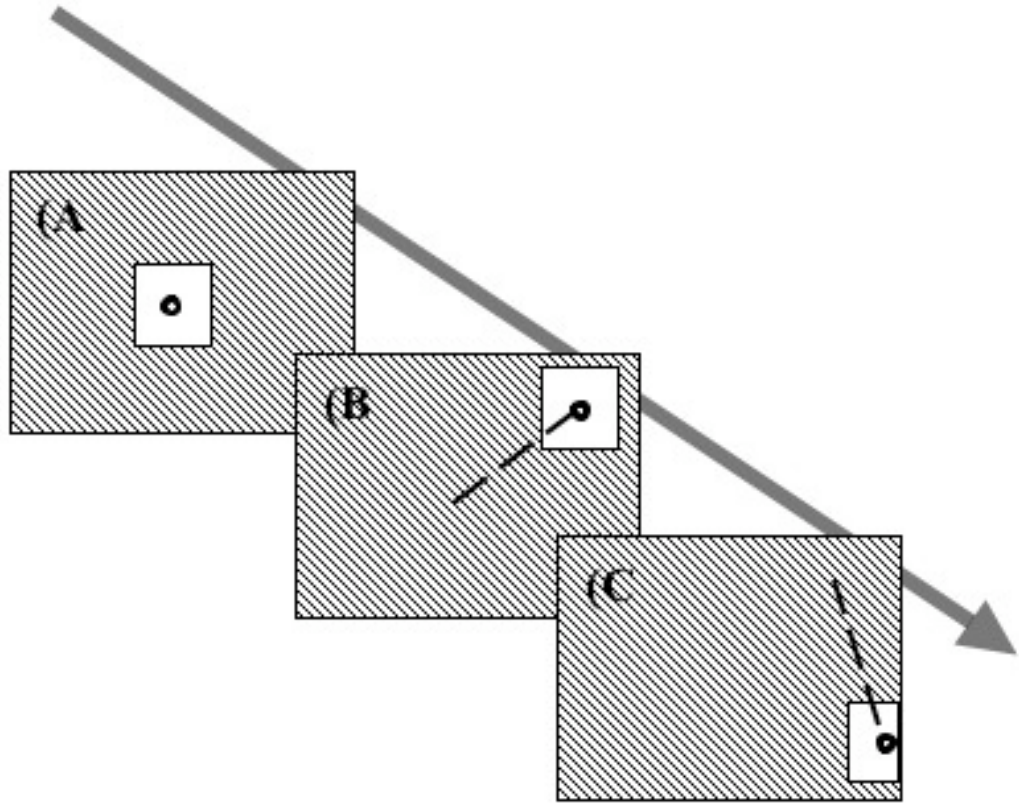


Figure 7.2. The gaze-contingent display used in Experiment 5. At each point in time, only a small segment of the image is fully visible (white square) whilst the background is distorted. (A. Fixation (black circle) starts in the centre. (B. A saccade (black line) is made to the top right of the display and detected by the eyetracker and the display changes. (C. Further fixations are followed closely by the "moving window". If fixation is near the edge of the display, only part of the window will be visible.

7.2.2 Results

The same measures were taken as those in Experiment 4. They are summarised in Tables 7.1 and 7.2 as a function of both target saliency and filtering type. The results were scrutinized using 3 x 3 mixed ANOVA, with Bonferroni corrected, post hoc *t* tests, in order to probe two main questions. Firstly, were salient regions reacted to quicker than non-salient regions, as in Experiment 4? Secondly, did this effect change according to different filtering conditions in the periphery? With these points in mind, a few key measures are then compared with Experiment 4 in order to explore the effect of gaze-contingent filtering.

		Filtering type								
		Blur			**Greyscale**			**Contrast**		
		S1	**S5**	**Cont**	**S1**	**S5**	**Cont**	**S1**	**S5**	**Cont**
Time to fixation (all trials) (ms) [1,2]	***M***	940	1087	1083	985	1138	1052	709	844	795
	SEM	*89*	*76*	*99.7*	*96.6*	*97.9*	*89.3*	*61.3*	*65.6*	*66.2*
Time to fixation (when cued) (ms) [1,2]	***M***	886	1053	1074	865	1184	1076	700	874	799
	SEM	*59.0*	*81.2*	*109.0*	*55.6*	*97.4*	*89.0*	*29.5*	*59.8*	*62.0*
Proportion of target absent trials where region is fixated [1,2]	***M***	0.46	0.42	0.30	0.45	0.43	0.31	0.34	0.39	0.23
	SEM	*0.033*	*0.025*	*0.025*	*0.028*	*0.022*	*0.031*	*0.023*	*0.022*	*0.019*

Table 7.1. Means and standard errors for the eye movement measures taken in Experiment 5(a). [1]Within-subjects effect of region type is statistically reliable at $p<.05$. [2]Between-subjects effect of filtering type is statistically reliable at $p<.05$.

Eye movement measures

Time to fixate the target was entered into a 2-way ANOVA. There was a reliable within-subjects effect of region saliency, $F(2,98) = 8.12$, $MSE = 35161$, $p = .001$. Pairwise comparisons showed that S1 regions were fixated reliably earlier ($M = 878$ ms) than either S5 or control regions ($M = 1023$ ms, $t(51) = 4.09$, $p = .001$ and $M = 976$ ms, $t(51) = 2.99$, $p<.05$ respectively). There was also a significant between-subjects effect of filtering type, $F(2, 49) = 4.07$, $MSE = 297593$, $p<.05$, which post hoc tests revealed to be due to a reliably lower time to target for the contrast manipulation ($M = 782$ms) than for the greyscale condition ($M = 1058$ms), $t(33) = 2.61$, $p<.05$. The comparison between contrast and blur ($M = 1036$ms) approached significance, $t(32) = 2.60$, $p = .069$, but the remaining comparison was not significant. There was no interaction between filtering type and region type, $F(4, 98)<1$.

An almost identical pattern was observed when looking at time to fixation in only those trials where the region had been cued. Region saliency had a significant effect, $F(2,98) = 13.17$, $MSE = 52000$, $p<.001$, and when inspected this was found to be caused by earlier fixation of S1 regions ($M = 816$ ms) than either S5 regions ($M = 1036$ ms, $t(51) = 5.7$, $p<.001$) or control regions ($M = 983$ ms, $t(51) = 3.7$, $p<.005$). The latter means did not differ reliably. There was a smaller effect of filtering, $F(2, 49) = 4.9$, $MSE = 191790$, $p<.05$. Post hoc tests showed that while blur led to longer times to fixation than the contrast condition, this was only marginally significant (Ms = 1004ms and 791ms, $t(32) = 2.7$, $p = .052$). There was, however, a reliable difference between the greyscale ($M = 1041$ms) and contrast conditions with the former leading to significantly longer times to target ($t(33) = 3.1$, $p<.02$). Filtering and region saliency did not interact, $F(4, 98)<1$. Thus the time to target was lower for S1 regions, this was true for all trials and for cued trials only, and this was not moderated by gaze contingent peripheral distortion.

As in Experiment 4 one can also ask how often regions captured attention when the target was not present. The saliency of the region had a large effect on how often it was fixated in target absent trials, $F(2,98) = 84.3$, $MSE = 0.0038$, $p<.001$. Both S1 and S5 regions were fixated on a greater proportion of trials than control regions (Ms = .42,

.42 and .28 respectively, both *t*s>9.2, *p*s<.001). The two types of salient region did not differ. There was also a significant effect of filtering type, *F(*2, 49) = 3.63, *MSE* = 0.0026, *p*<.05. Pairwise comparisons showed that there was no difference between the blur (*M* = .40) and grey (*M* = .40) conditions. Participants in the contrast condition (*M* = .32) looked at the pre-defined regions less often than those in either the blur or contrast conditions, but these differences did not reach the criterion for significance. The interaction between region saliency and filtering type was not reliable, *F(*4, 98) = 2.16, *MSE* = 0.0038, *p* = .079.

		Filtering type								
		Blur			**Greyscale**			**Contrast**		
		S1	**S5**	**Cont**	**S1**	**S5**	**Cont**	**S1**	**S5**	**Cont**
Hit rate	***M***	0.875	0.894	0.824	0.82	0.79	0.8	0.86	0.91	0.87
	SEM	*0.032*	*0.028*	*0.034*	*0.027*	*0.034*	*0.042*	*0.03*	*0.018*	*0.024*
***d* prime [2]**	***M***	2.52	2.72	2.04	1.8	1.84	1.6	2.39	2.33	2.46
	SEM	*0.24*	*0.22*	*0.23*	*0.219*	*0.21*	*0.256*	*0.233*	*0.253*	*0.14*
Reaction time (ms) [1,2]	***M***	2780	3332	3525	2999	3758	3565	2092	2497	2402
	SEM	*229*	*272*	*269*	*248*	*341*	*293*	*147*	*181*	*183*

Table 7.2. Means and standard errors for the response measures taken in Experiment 5(a). [1]Within-subjects effect of region type is statistically reliable at *p*<.05. [2]Between-subjects effect of filtering type is statistically reliable at *p*<.05.

Response measures

Statistical analysis of the hit rates in each condition showed no effect of region saliency, $F(2,98) = 2.07$, $MSE = 0.00813$, $p = .132$; filtering condition, $F(2,49) = 2.48$, $MSE = 0.033$, $p = .094$; or their interaction, $F(4,98) = 1.37$, $MSE = 0.00813$, $p = .25$. Region saliency did not affect *d* prime, $F(2,98) = 1.94$, $MSE = 0.520$, $p = .150$. However, there was a reliable effect of filtering condition on sensitivity, $F(2,49) = 4.68$, $MSE = 1.65$, $p<.05$, and this was due to a lower *d* prime for participants in the greyscale condition ($M = 1.75$) than for either those in the blur condition ($M = 2.42$) or those in the contrast condition ($M = 2.39$; both ts = 2.5, ps<.05). There was no interaction between saliency and filtering, $F(4,98) = 1.41$, $MSE = 0.520$, $p = .24$.

Correct reaction time to target present trials was affected by both target region saliency, $F(2,98) = 27.45$, $MSE = 195453$, $p<.001$, and filtering condition, $F(2,49) = 5.94$, $MSE = 3000397$, $p = .005$, although these factors did not interact, $F(4,98) = 1.76$, $MSE = 195453$, $p = .142$. In terms of region saliency, participants were faster to respond when searching for an S1 region than for either S5 or control regions (Ms = 2624, 3195 and 3163ms, respectively, both ts>5, ps<.001). S5 and control regions were not significantly different. In terms of filtering type, the contrast manipulation ($M = 2330$ms) led to significantly shorter reaction times than either blur ($M = 3212$ms, $t(32) = 2.9$, $p<.05$) or greyscale ($M = 3441$ms, $t(33) = 3.4$, $p<.01$).

Comparisons with Experiment 4

There were widespread effects of saliency on both eye movement measures and manual reaction time. There were also some differences between the different filtering conditions. In order to clarify these, data for some of the measures were compared with those from Experiment 4. What was the effect of filtering, in comparison to the standard, undistorted search? In each case data from the standard search task was entered as an additional level of the filtering condition factor, giving a 4 (one standard and 3 gaze-contingent viewing conditions) by 3 (types of target region) mixed ANOVA.

The time to fixate the target, for correct target present trials when cued, is shown for all levels in Figure 7.3. It is clear from this figure that there is a robust effect of saliency across viewing conditions. This effect is statistically reliable, $F(2,128) = 13.03$,

MSE = 49683, $p<.001$. Across these conditions S1 regions were fixated after an average of 783 ms, significantly earlier than both S5 ($M = 972$ ms, $t(67) = 5.5$, $p<.005$) and control regions ($M = 923$, $t(67) = 3.8$, $p<.001$). There was no reliable difference between S5 and control regions. A non-significant interaction between region saliency and filtering, $F(6,128)<1$, indicates that this general pattern was unchanged by the peripheral filtering. On the other hand, filtering had a strong main effect on the time to fixate the target, $F(3,64) = 7.54$, *MSE* = 156861, $p<.001$, independent of saliency. Unsurprisingly, search without any distortion led to earlier target region fixation (after a mean of 735 ms) than the blur condition ($M = 1004$ ms, $t(31) = 3.5$, $p<.01$) or the greyscale condition ($M = 1041$ ms, $t(32) = 3.9$, $p<.005$). The contrast condition ($M = 791$ ms) was not reliably impeded compared to normal search and had significantly faster times to target than the greyscale condition ($t(33) = 3.1$, $p<.05$) or the blur condition ($t(32) = 2.7$, $p = .05$). No other paired comparisons were reliable.

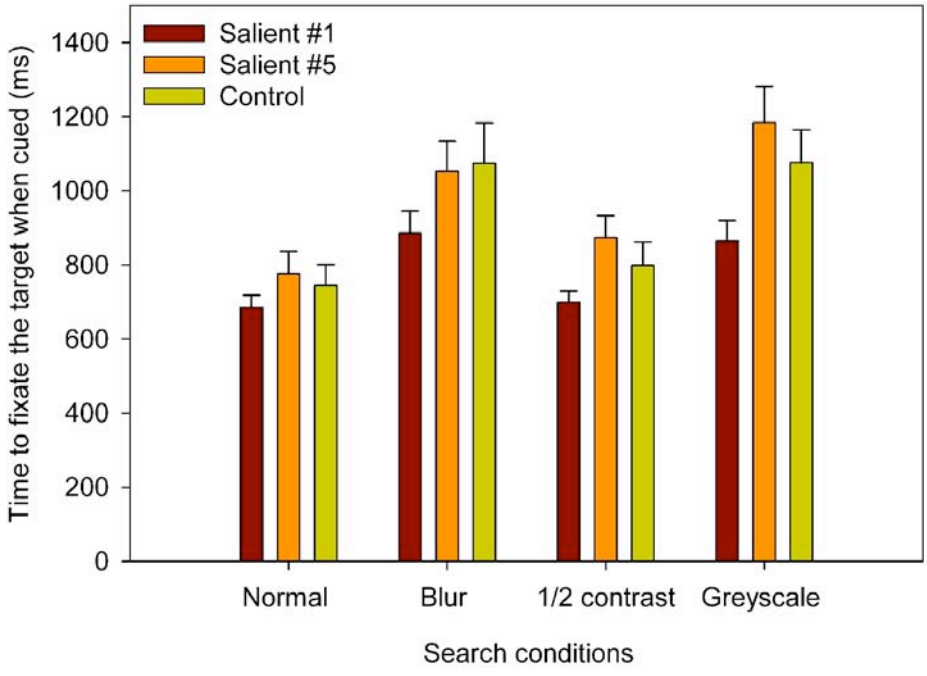


Figure 7.3. The mean time to fixate the target region in Experiments 4 and 5(a), as a function of region salience and viewing conditions. Error bars show one standard error of the mean.

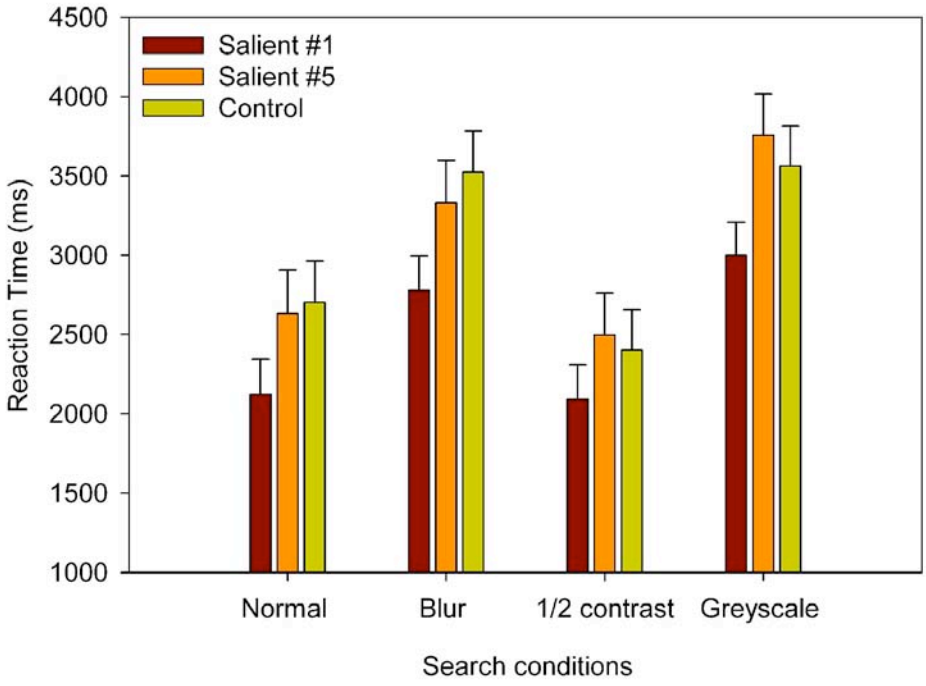


Figure 7.4. Mean reaction time (with standard error bars) across the different region types and search conditions in Experiments 4 and 5.

A highly similar pattern is seen in the reaction time data (Figure 7.4), with the effect of saliency being even more pronounced: $F(2,128) = 37.43$, $MSE = 184782$, $p<.001$. Participants responded faster to S1 regions ($M = 2498$ ms) than either S5 ($M = 3055$ ms) or control regions ($M = 3048$ ms, both ts$(67)>7$, ps$<.001$). S5 and control regions did not differ significantly. The search conditions affected reaction time, $F(3,64) = 5.54$, $MSE = 2736761$, $p<.005$. Search with greyscale filtering ($M = 3441$ ms) was slower than both the manipulated contrast condition ($M = 2330$ ms, $t(33) = 3.4$, $p<.01$) and undistorted search ($M = 2486$ ms, $t(32) = 2.7$, $p<.05$). The comparison between the blur ($M = 3212$) and contrast conditions was marginally significant ($t(32) = 2.9$, $p = .054$) and none of the other comparisons were reliable. Thus search times in the contrast condition were equivalent to those when no gaze-contingent filtering took place. The interaction between filtering and saliency was not reliable, $F(6,128) = 1.30$, $MSE = 184782$, $p = .260$

A further point of interest concerns the processing allocated to different regions. In Experiment 4 it was suggested that longer reaction times might not be due to time to

reach the target but rather time to encode or process its contents. In order to gauge this the first fixation duration on the region was compared across all 4 viewing conditions, regardless of which region was the target. This measure is thought to reflect the amount of processing of foveal information (Rayner, 1998). As such peripheral distortion might not be expected to have any effect (the information present at the centre of fixation was the same in all conditions). Other research has reported some effects of peripheral masking on current fixation duration (Greene, 2006; Loschky & McConkie, 2002; van Diepen & d'Ydewalle, 2003). Presumably part of the current fixation time is spent planning a saccade to a peripheral location, and if the decision regarding where to go is confounded by more noise, due to distortion and more competition, fixation duration might be lengthened. The mean first fixation duration on a region is shown in Table 7.3.

There was a significant effect of region saliency on first fixation duration, $F(2,128) = 13.16$, $MSE = 748$, $p<.001$. Control regions ($M = 268$ms) were associated with shorter first fixation durations than both S1 ($M = 293$ms, $t(67) = 5.1$, $p<.001$) and S5 ($M = 281$ms, $t(67) = 2.4$, marginally significant at $p = .058$). S1 regions were fixated for around 25 ms longer, on average, than control regions and they were also fixated for significantly longer than S5 regions ($t(67) = 2.8$, $p<.05$).

Viewing conditions		**Region type**		
		S1	**S5**	**Control**
Normal search (Exp. 4)	***M***	268	259	245
	SEM	*7.1*	*8.6*	*10.3*
Blur	***M***	273	260	266
	SEM	*10.1*	*9.1*	*7.9*
Greyscale	***M***	322	314	284
	SEM	*14.2*	*12.6*	*12.9*
Contrast	***M***	307	291	279
	SEM	*15.2*	*10.7*	*10.2*

Table 7.3. Mean first fixation duration (in ms) on each type of region in the different viewing conditions. Standard errors are shown in parentheses.

Viewing conditions also had an effect on first fixation duration, $F(3,64) = 5.39$, $MSE = 5009$, $p<.005$. This was mainly due to elevated first fixation durations in the

greyscale condition which were around 50 ms longer than those in the standard search condition ($t(32) = 3.5$, $p = .005$) and which were also longer than the blur condition ($t(33) = 2.9$, $p<.05$). No other comparisons were reliable. Thus, relative to undistorted search which showed the shortest first fixation durations, those in the blur and contrast conditions were less impaired than those in the greyscale condition. Viewing conditions did not interact with region saliency, $F(6,128) = 1.56$, $MSE = 748$, $p = .164$.

Model predictions for filtered images

Up until now I have assumed that the peripheral filtering used would change the predictions of the model and therefore the relative saliency of scene regions. Given that the gaze contingent conditions led to the same effect of saliency as that seen in normal search, it is important to ask to what degree the filtering actually affected the model predictions. The three types of filtering would affect the saliency map predictions in different ways. For example, changing the contrast across the whole image might not actually change the predictions at all, as the relative, local differences would remain the same. To get an idea of the effect of the filtering manipulations on the saliency model, I used it to process the filtered images. Of course, in the actual stimuli these would have included an intact window at fixation, but here I will focus on how the peripheral saliency map changes.

Figure 7.5 shows some examples, and it is clear that in some cases the saliency ranking of different regions is altered. On the other hand, the same regions often continue to be salient despite filtering. For the purposes of the current experiment what is of interest is the degree to which the first (and fifth) most salient regions continued to be salient. In order to get a quantitative estimate of the effect of the filtering manipulations on the saliency map for each scene, I compared the saliency map from the normal image with those from the three different filtering conditions. Three, two-dimensional correlations were performed (see Appendix B for formula). The saliency maps were $1/16^{th}$ the scale of the original image. If the saliency map were identical before and after the filtering manipulations, then the correlation coefficient would be 1. Zero correlation would indicate that the salient points changed completely, whilst a negative coefficient

would suggest that points with low saliency tended to have higher saliency after filtering and vice versa.

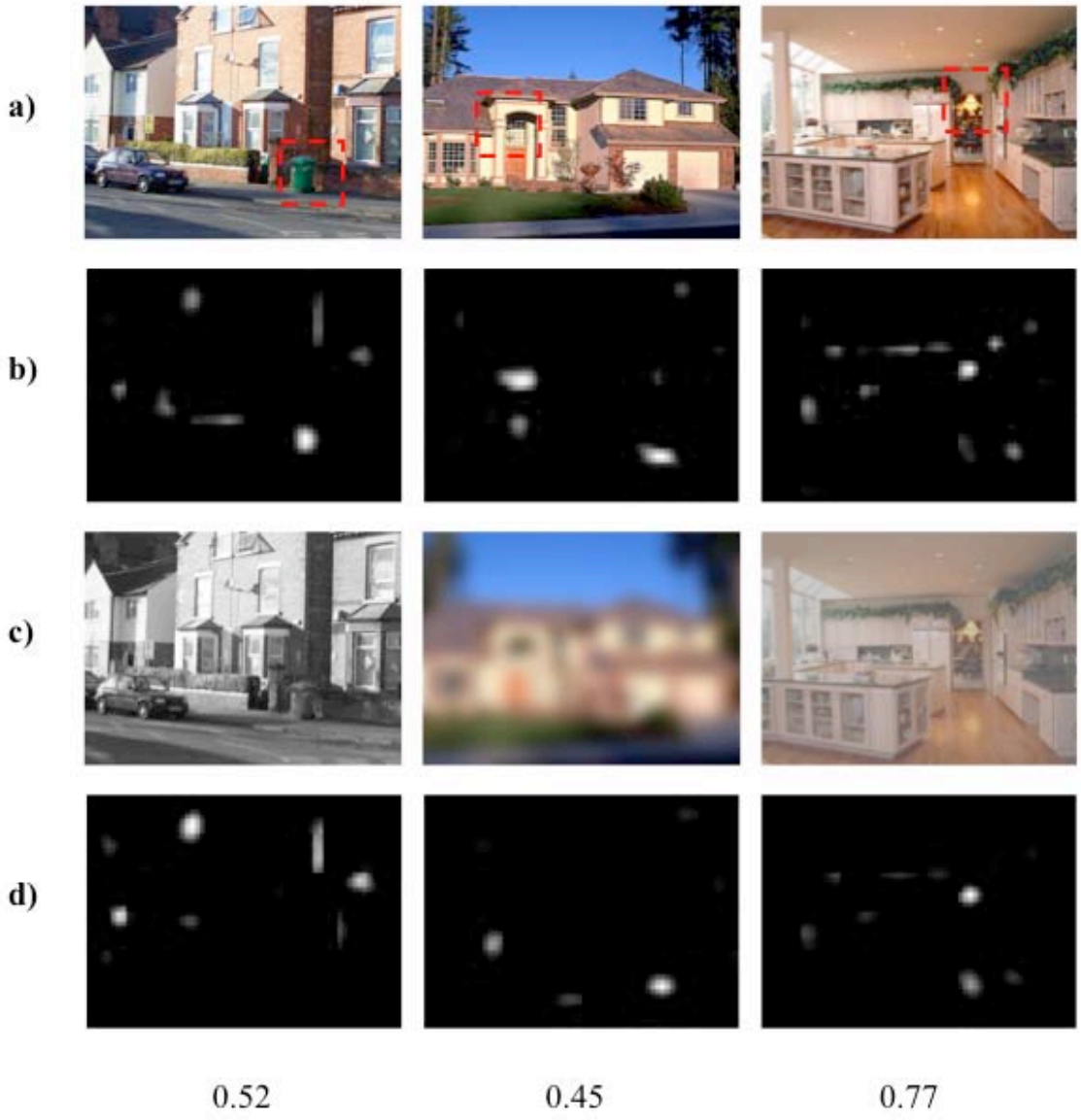


Figure 7.5. The effects of filtering on three example stimuli. The figure shows the original images (a), with the most salient, S1 region highlighted, and saliency maps for each (b). The three filtering conditions are illustrated in c): greyscale (left column), blur (middle) and half contrast (right). These manipulations change the resulting saliency maps (d). For example, in the left column a green dustbin that was highly salient no longer stands out. The 2-dimensional correlation coefficient measuring the change between the saliency maps is shown below the figure, for each stimulus.

Across all stimuli, the mean, bivariate correlations with the normal image were 0.58 (*SEM* = 0.021), 0.86 (*SEM* = 0.012) and 0.94 (*SEM* = 0.007) for the blur, greyscale and contrast stimuli respectively. These values suggest two things. First, none of the

correlations are particularly low (or negative) which means that the filtering manipulations did leave much of the saliency information intact. A manipulation that led to totally different predictions might be more useful for subsequent research, perhaps one that filtered only part of the image. Second, there was variation both between and within conditions in the degree to which the saliency map changed. Lowering the contrast across the image hardly modified the saliency map at all, and hence the correlation is close to 1. Removing the colour had slightly more effect, and blurring the images made the most difference. Given these observations, one might expect the saliency effect to be weakest in the blurred, gaze contingent condition, as it is here that the saliency of scene regions changes the most and the S1 regions were less likely to be salient in comparison to other parts of the scene. However, as the previous sections demonstrated, there was no interaction between saliency and filtering type.

Did the degree of change in the saliency map make a difference to the key finding that salient target regions were fixated earlier? To test this, I performed an additional, post hoc analysis. Within the blur condition, the stimuli were grouped into thirds based on the correlations above. The results from the third with the highest correlation (i.e. those where the saliency map was similar to that from the normal image) were compared to the third with the lowest correlation (where saliency changed the most). When the condition was subdivided in this way, the mean correlation between the filtered and normal maps was 0.77 for the highest third and 0.33 for the lowest. This was most appropriate for the blur condition as there was much more of a variation between the different scenes, and I concentrated on the key result of the difference between the time to fixate the target when cued, for S1 versus control regions. A 2 x 2 repeated-measures ANOVA tested the effects of region (S1 v. control) and saliency map correlation with normal (high v. low). If the change in saliency after filtering makes a difference then one would expect an effect of region only when there was a high correlation (because S1 and control regions would still be different) and not when filtering changed the saliency map (because S1 regions were less likely to stand out).

The main effect of saliency map correlation was not reliable, $F(1,16) = 2.06$, $MSE = 288413$, $p = .17$. There continued to be a reliable effect of region saliency,

$F(1,16) = 7.65$, $MSE = 192930$, $p<.05$, but this was qualified by a significant interaction $F(1,16) = 4.92$, $MSE = 218581$ $p<.05$. Looking at the marginal means, there was only a slight difference between the time to fixate the regions in stimuli with a low correlation (S1, $M = 1107$ ms, $SEM = 154$ms; Control, $M = 1151$ms. $SEM = 130$ms). This difference was much greater in the high correlation condition (S1, $M = 1043$ms, $SEM = 118$ms; Control, $M = 1589$ms, $SEM = 179$ms). To summarise this analysis, therefore, the filtering manipulation did not always change the model's predictions, but when it did the advantage for salient over control regions was removed.

7.2.3 Discussion

Experiment 5(a) showed reliable effects of saliency on eye movements and task performance, and this confirms what was found in Experiment 4 in the previous chapter. In all filtering conditions there was evidence that attention was allocated towards the salient regions earlier. This was particularly true for the most salient region in the scene (S1) and data for a point selected by the model later (S5) was more noisy, often being indistinguishable from the random control regions. Participants were faster to confirm the presence of a salient region. This was a large and robust effect, which shows that the saliency map model can make an experimental difference to task performance by identifying regions that will be found and processed more efficiently than chance. Although accuracy was high and did not differ reliably, there was a trend in the predicted direction such that sensitivity to control regions was generally lower. There is evidence that attention is preferentially allocated to the centre of a display (Tatler et al., 2005) and that short saccades are favoured in natural scenes, so it is important to stress that the difference between the regions cannot be explained by their eccentricity. Salient regions were actually slightly further from the centre than control regions, but were still fixated earlier.

As discussed previously these results show that saliency can make a difference in search and that it is not totally overridden by knowledge of a target. There was also important converging evidence in terms of the target-absent trials. In these cases, if eye movements were guided solely in terms of target features then the different regions would be fixated rarely and equally often, as none would be a good match with the target.

However, in both Experiments 4 and 5(a) the regions were fixated and this was more likely for salient regions.

Importantly, peripheral filtering using a gaze-contingent display did not remove the benefit found for salient regions as targets. This is a problem for the saliency map model as it suggests that saliency does not affect the guidance of long-range saccades into the periphery. I will discuss the implications of this result in more detail at the end of this chapter. To summarize, although filtering made search harder, the trend was exactly the same as in standard search. This might mean that the filtering manipulations were not such that they changed the predictions of the model. Mannan et al. (1995) reported that people looked in roughly the same places when images were transformed by high- and low-pass filtering. Here, there was similarly little effect of global image manipulations on fixation behaviour with regard to the regions of interest.

An explanation for the lack of interaction with peripheral filtering which is more problematic for the saliency model is that the benefit for salient regions is not completely due to bottom-up selection. By this account the model is identifying areas which are easier to find, not because of reflexive, feature-based attentional selection but because they are easier to categorise and richer in meaning. For example, people may have looked at these regions earlier because they are primed to look at objects that are semantically informative in the context of the scene, and not visually salient per se. Alternatively, given that the experiments were search tasks, the salient regions may have contained more information that would allow participants to predict where they were in advance, based on knowledge of objects, and so guide the eyes top-down. Looking at the image patches used, it is certainly true that control regions were more likely to contain blank areas of wall and sky and less likely to contain objects and meaningful items. In the present experiment, the effect of saliency on fixation duration is consistent with the fact that salient regions were more meaningful.

It is possible that the saliency advantage in search could be caused by top-down knowledge of where regions are likely to appear, with salient regions more likely to be linked to such knowledge (it is easier to predict where a chair might be than a piece of wall). However, in the memory task in Experiment 3 no such information was given in

advance and a saliency advantage was still seen. In order to clarify this issue in a search task, two further experiments were carried out varying the content of the scene. First, in a manipulation that was expected to affect the semantics of the scene, the display was inverted.

7.3 Experiment 5(b): searching in an inverted scene

7.3.1 Introduction

This experiment repeated the search procedure but modified half the stimuli by inverting them. This manipulation is useful as it preserves all local visual information. The saliency map, and the predictions of any purely bottom-up model, will be exactly the same (although the spatial position of salient regions will be flipped). On the other hand inversion effects have been reported in some classes of stimuli, affecting the way in which people process features. In face recognition, for example, different responses to upright and inverted face stimuli are used as evidence for different processing styles—holistic or analytic (Valentine, 1988). This may or may not be reflected in different eye movement patterns (Henderson, Williams, & Falk, 2005). In other classes of stimuli, people are less accurate and slower to discriminate upside-down items when they have been trained on upright exemplars and this is true for novel stimuli (Greebles; Gauthier & Tarr, 1997) and pictures of houses and textures (Husk, Bennett, & Sekuler, 2007). This may be because configural processing (based on the spatial layout of features) is disrupted, or simply due to practice with the canonical orientation.

In change detection experiments scene inversion has been used as a control to reduce the contextual guidance that is assumed to draw attention to "central" changes quicker than to those of marginal interest (Kelley, Chun, & Chua, 2003). In Kelley et al.'s experiment (which attempted to control the visual saliency of changed items) the central-peripheral difference was removed for inverted scenes. Little research has specifically examined what inverting scenes does to gist acquisition and eye movements, but it seems likely that this will impair the perception of spatial layout and predictions of where expected objects or features will appear. If the advantage for salient regions in search is because of their bottom-up conspicuity and not their semantic or contextual meaning then it should be unaffected by inversion. If inversion changes the effect it

would indicate a role of scene semantics over and above visual saliency. If inverting the scene makes the information at fixation more difficult to process, one would also predict longer fixation durations.

7.3.2 Method

Participants

A new group of 17 student participants were recruited and they all had normal or corrected-to-normal vision and achieved good calibration.

Stimuli, design and apparatus

The stimuli and design for this experiment replicated those of the previous region search experiments. In order to modify the top-down semantic context, half of the stimuli were inverted. To give an equal number of pictures with each type of target region the total number of target present stimuli was increased slightly to 48 (with 24 inverted). S1, S5 and control regions were equally represented in each inversion condition. For inverted scenes the target region was also inverted, so no additional steps were required to mentally rotate the target. Figure 7.6 shows an example of each type of stimuli. Half of the target-absent pictures were also inverted so that picture orientation did not provide any information relevant for the search response.

Figure 7.6. Two of the stimuli used in the experiment. Both show outdoor scenes, but one is from the inverted condition (left) whilst the other is a non-inverted control.

Procedure

The procedure was exactly the same as the standard search task from Experiment 4. A gaze contingent display was not used and upright and inverted pictures were interspersed in a random order.

7.3.3 Results

In order to compare the results with Experiments 4 and 5(a), the correct target-present reaction time, the time to fixate the target on those trials where it was cued, and the probability of each region being fixated in target-absent trials were computed. The duration of the first fixation on the target region was also analysed. The subject means were subjected to a within-subjects ANOVA with factors of scene orientation (upright vs. inverted) x target region (S1, S5 or control).

Reaction time

The mean RT across all types of trial was 2354ms, comparable to that in Experiment 4, and accuracy was relatively high (82% hits). There was a reliable main effect of region type on the correct target present RT, $F(2,32) = 5.92$, $MSE = 214479$, $p<.01$. Pairwise comparisons showed a significant difference between S1 and control regions ($t(16) = 3.27$, $p = .014$) but no other differences. S1 targets resulted in quicker mean responses (2168ms) when compared with control regions (2554ms) while S5 regions were intermediate between the two (2341ms).

Although the mean RT to inverted regions was slightly longer than to upright images (2427ms and 2281ms respectively) this difference was not reliable, $F(1,16) = 1.64$, $p = .22$, and neither was the interaction, $F(2,32)<1$.

Eye movement measures

There were very few effects on the time to first fixate the different regions in this experiment. Figure 7.7 depicts the mean time before fixating the target when it was cued as a function of region saliency and scene orientation. The pattern is qualitatively different between the two levels of inversion, with that in the normal, non-inverted trials appearing more similar to the previous experiments than in the inverted condition. There was a marginal effect of inversion, $F(1,16) = 4.46$. $MSE = 96237$, $p = .051$, although this

was in a counterintuitive direction, with inverted scenes leading to slightly quicker search times overall (*M*s = 727 and 857ms respectively). There was no effect of region saliency and no interaction, both $F(2,32)<1$.

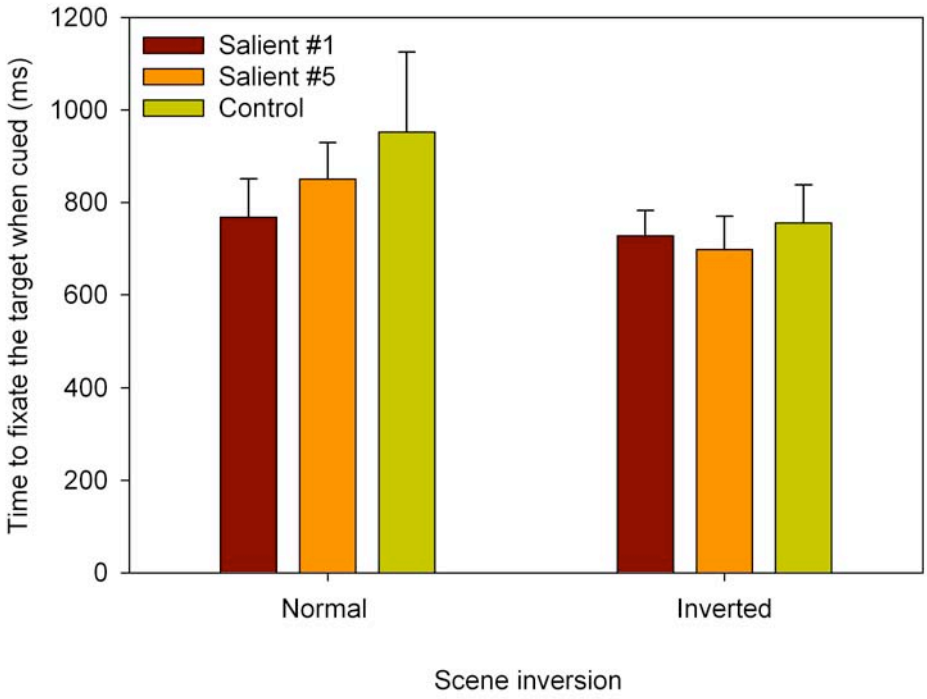


Figure 7.7. The mean time to fixate the target region, when searching for it in normal and inverted scenes. Error bars show plus one standard error of the mean.

The previous analysis suggested that the inversion manipulation did not lead to an advantage for salient regions over controls, which was the pattern seen in all the conditions in Experiment 5(a). Looking at the proportion of target absent trials where each type of region was fixated, inversion had a reliable effect, $F(1, 16) = 13.77$, $MSE = 0.003$, $p<.005$. Inverted regions of all types tended to be looked at more frequently (in 0.36 of trials on average, $SEM = 0.02$) than those shown the normal way up ($M = 0.32$, $SEM = 0.02$). There was also a large effect of region saliency, $F(2,32) = 67.4$, $MSE = 0.003$, $p<.0001$, though the interaction was not reliable, $F(2,32) = 2.38$, $MSE = 0.006$, $p = .11$. Both S1 ($M = 0.36$, $SEM = 0.02$) and S5 regions ($M = 0.40$, $SEM = 0.02$) were fixated more often than control regions ($M = 0.26$, $SEM = 0.03$; ts(16) = 7.69 and 10.95 respectively, both ps<.001). The difference between the two types of salient regions was

also reliable, and it was in the opposite direction to that predicted by the model, with the fifth most salient region being fixated more often than the first ($t(16) = 3.52$, $p<.01$).

How did the first fixation duration on the region compare with previous experiments? The mean of this measure was 252ms overall, which is very similar to that reported previously. There was no effect of inversion on this duration (normal: M = 254ms, SEM = 10ms; inverted: M = 249ms, SEM = 8; $F(1, 16)<1$). However, as in previous experiments there was an effect of region saliency, $F(2,32) = 3.66$, $MSE = 751$, $p = .037$), and comparisons indicated that S1 regions (M = 261ms, SEM = 10) were fixated for longer than both S5 (M = 243ms, SEM = 7) and control regions (M = 251ms, SEM = 10), though this was only reliable in the former ($t(16) = 3.23$, $p<.05$). The S1 and S5 conditions did not differ, and there was no interaction, $F(2,32) = 1.55$, $MSE = 923$, $p = .23$.

7.3.4 Conclusions

The results from this experiment were somewhat inconclusive. As in previous experiments there was still an effect of saliency on manual reaction time, and this was unaffected by inverting the scene. However, the key finding that the most salient region was fixated earlier than controls was not supported by the present results. This may have been due to inverting the scenes and the interaction, though not significant, was in the predicted direction. There are some problems interpreting this effect. One is that this experiment had less power than Experiments 4 and 5(a) (because the trials were split into inverted and non-inverted conditions) and this made finding a significant result less likely. Another is that in target absent trials there was still a benefit for salient over control regions, as found previously. This might be because a different gaze selection strategy is used in the absence of the target signal, perhaps one that is less affected by scene inversion than in target present trials. Finally, contrary to predictions, inversion did not affect the duration of the first fixation on a region. Salient regions continued to be fixated for longer than control regions, and this did not interact with the flipping of the scene.

Taken together, these results suggest that part of the benefit for salient regions in this task is modified by a change in the global layout, even though this change would

leave bottom-up saliency unchanged. On the other hand, there were still some effects of saliency so this is unlikely to be the full explanation. To further clarify the influence of the meaning of the region, a final experiment was carried out using a gaze contingent display where all peripheral information was masked.

7.4 Experiment 5(c): searching in a fully-masked display

7.4.1 Introduction

The findings from Experiment 5(b) suggest that at least some of the advantage for salient regions is due to their meaningfulness and not their ability to attract attention in the periphery. If this is the case then it is not surprising that the gaze contingent manipulations failed to reduce this effect; manipulating the peripheral information available did not change the targets' meaning. A more severe test of this possibility is to completely mask the scene beyond fixation. In this case there are no features from which to form a saliency map, other than those within the window. How will search proceed in this case? If certain regions are found quicker with no peripheral information then this must be due to the viewer's knowledge of where particular objects or features are likely to occur, either generally within the boundaries of the image (e.g. chimneys are likely to be near the top of the scene) or in relation to the features at fixation (e.g. chimneys are likely to be above the door).

7.4.2 Method

Participants

A new group of 17 participants with the same general characteristics as those from previous experiments took part. None had completed other experiments with these stimuli.

Stimuli, design and apparatus

The pictures, design and eye tracking apparatus were the same as those used in the other experiments in this chapter. The gaze-contingent display was programmed using a square region of the unmodified image (with the same dimensions as previously) as the foreground, and this window followed fixation. The background was a uniform grey

image, chosen to minimise the luminance difference between the window and the mask. All pictures were shown upright.

Procedure

The procedure was exactly the same as that in the gaze contingent search task, Experiment 5(a). During any one fixation, only the window around the current fixation location was visible, so participants had no peripheral, visual cues to guide the eyes. Instead, they had to base their eye movements on their expectations of where the region would be. An extra practice session and regular recalibrations ensured that the match between fixation position and the visible window was optimal.

7.4.3 Results

Whilst accuracy was maintained to a similar degree as previously (87% hits), the correct target present RT indicated that the fully-masked display made the task much more difficult. Across all target types, participants took 5625ms to make their manual response, and this was much longer than in Experiments 4, 5(a) and 5(b) (independent *t* tests collapsed across all conditions, all *t*s>5, all *p*s<.001). The measures taken are summarised in Table 7.4 for the different types of region. I began by analysing RT.

Response measures

Interestingly, there was still a reliable effect of region type on the RT in this experiment, $F(2,32) = 3.61$, $MSE = 1236311$, $p<.05$. However, this result was somewhat smaller than that seen in previous experiments, and there was more variability in the speed at which participants responded. S1 regions were responded to quickest. Paired comparisons showed that whilst the most salient region led to faster responses than the S5 regions ($t(16) = 3.19$, $p<.01$), neither S1 nor S5 were reliably faster than control regions.

		Region type		
		S1	**S5**	**Control**
Correct target present RT (ms)	***M***	5093	6116	5667
	SEM	*458*	*434*	*506*
Time to first fixate the target when cued (ms)	***M***	2094	2357	1826
	SEM	*121*	*203*	*141*
Proportion of target absent trials where fixated	***M***	0.67	0.65	0.54
	SEM	*0.03*	*0.03*	*0.03*
First fixation duration on region (ms)	***M***	283	286	291
	SEM	*12*	*14*	*11*

Table 7.4. The mean and standard error of each of the measures taken in Experiment 5(c).

Eye movement measures

Region type had a reliable effect on the time taken to move to the target region, $F(2,32) = 3.39$, $MSE = 353841$, $p < .05$) However, comparing between the levels it is clear that the effect here is different from that in Experiments 4 and 5(a). Importantly, control regions were actually found slightly faster than salient regions. Only the comparison between S5 and control regions was reliable ($t(16) = 2.57$, $p<.05$). All other differences failed to reach significance.

As was the case in previous experiments, the proportion of target absent trials where the region was fixated at least once was lower for control regions than the salient areas. This effect was reliable, $F(2,32) = 21.61$, $MSE = 0.001$, $p<.001$, and pairwise comparisons confirmed that control regions were fixated less often than either of the salient regions (both ts$(16)>4.5$, ps$<.001$). S1 and S5 regions did not differ.

Elsewhere in this chapter I have pointed to differences in the first fixation duration as indicative of both the processing allocated to a region (and therefore its semantic informativeness) and possibly the planning of the next saccade (as the duration

was prolonged when gaze contingent filtering was applied to the periphery). Interestingly, although in all previous search experiments salient regions were fixated for longer, in the present experiment there was no effect of region type on the first fixation duration, $F(2,32)<1$. The mean first fixation duration across all regions was prolonged relative to the standard search task in Experiment 4 ($t(31) = 2.2$, $p<.05$).

7.4.4 Conclusions

This experiment was a useful control for differentiating between knowledge-based eye guidance (which depended on the information in the target region and at fixation) and image-based eye guidance (which depended on peripheral information such as saliency). In this task, the benefit for salient regions in target present trials was removed. The most salient region was not responded to quicker than a random control, and the control was actually fixated earlier. This is a reversal of the effect seen elsewhere. It suggests that salient regions are fixated early, not just because people know where they are likely to occur, but because of peripheral, image-based saliency. By the same logic, the continued effect of saliency on the probability of fixating a region in target absent trials is hard to explain. This cannot be just to do with salient features attracting attention from outside the foveal and parafoveal span, because in the present experiment this area was empty. The analysis of first fixation duration was also interesting. This measure was prolonged by the peripheral masking, but it was unaffected by region saliency.

7.5 General discussion

The three experiments in this chapter qualify some of the findings in Chapter 6. In that chapter, salient regions were found quicker than non-salient regions, and the manipulations used here tried to distinguish between the factors that contributed towards this effect. In Experiment 5(a) there continued to be a trend showing that salient regions were fixated, and responded to, earlier, but this was moderated by scene inversion (Experiment 5(b)) and completely masking the periphery (Experiment 5(c)).

7.5.1 Explaining the advantage for salient regions

Why was attention allocated preferentially towards salient regions? I will continue to consider this in terms of a distinction between image-based and knowledge-based

guidance. The saliency map model explains the saliency advantage in terms of saccade target selection. Possible fixation locations are represented in an explicit saliency map, and higher peaks will represent salient regions. In competition with other salient non-target features, control regions will be less likely to "win" and so will be fixated later than salient regions. In an efficient search task, the saliency of distracter regions might be reduced based on target features so all targets might be expected to be fixated quickly, irrespective of saliency (as in Experiment 1(a)). However, the present findings suggest that this is not the case as there was a significant difference between the time taken to move to a salient versus a control target region. This may be because searching for a region is a less practised and less efficient task so some non-target regions were not filtered out and remained to compete with the control regions. Alternatively, it may be that even in a weighted saliency map with one winning target, a higher underlying peak is quicker to evolve or reach a threshold than a smaller, less salient one.

If a bottom-up saliency map is used to represent peripheral targets and control their selection then filtering information in the periphery would be expected to have an effect. In fact, in Experiment 5(a) the saliency benefit was present to the same degree as in the normal search. The gaze-contingent filtering did impede search, but filtering did not modify the effect of saliency. There are several possible reasons for this. Firstly, the filtering used may have left enough information intact for an unchanged saliency computation. While the blurring function used was quite severe, perhaps at higher levels of blur or distortion in other ways might ameliorate the saliency benefit. It is possible that the size of the window may have been too large. Both van Diepen et al (2001) and Greene (2006) used smaller windows and this would restrict the information available outside the fovea even more. However as the saliency model is designed to predict large shifts of attention (including. those of amplitude greater than 3˚ which would take them outside the window) this is not a satisfactory explanation. Loschky and McConkie (2002) continued to find effects of peripheral filtering even at their largest window size (a radius of 4.1˚ which is bigger than the window used here). It may be that participants only moved to the regions of interest from adjacent areas, in which case they may have identified them when within the window where the visual information was not degraded.

If the same regions that are salient in a normal image remain capable of preferentially attracting attention when colour or high spatial frequency information are removed then it is surprising. Further model evaluations should be made, comparing model predictions in normal and filtered images. If there is a certain amount of redundancy involved in the saliency computation in natural scenes then it may be that places that have high orientation contrast, for example, also have high colour and intensity contrast. In this case the saliency map for a blurred or greyscale image might not change that much. A preliminary analysis compared the extent to which the saliency map changed after filtering to the size of the saliency advantage. The results were promising and showed that, when the saliency map changed a lot, different regions were selected and so the advantage for salient regions was lost. Further experiments where both colour and high spatial frequency information is removed might remove the difference between salient and control regions. Another improvement would be to selectively filter only part of the image, say half, balanced to sometimes be the half that contained the target and sometimes not.

An alternative, knowledge-based possibility—that the benefit for salient regions came about because these regions were more informative—can be evaluated by considering the processing allocated to a region. It has been assumed in most research that the duration of fixations, and particularly the first fixation, reflects the amount of processing on an object, and perhaps how informative or interesting it is. For example, incongruous objects in scenes tend to be fixated for longer (Henderson et al., 1999; Loftus & Mackworth, 1978), In the present experiments, and across almost all of the search conditions, the first fixation duration on salient objects was higher. This is consistent with the salient regions being more interesting, more informative, or perhaps being more likely to contain recognisable objects. Could this difference also explain the finding that salient regions were fixated earlier?

7.5.2 Semantic differences between the regions

It is difficult to quantify the meaning of scene regions. Few people have quantified or controlled for the semantic value of to-be-fixated areas. Henderson et al. (2007) showed participants small parts of a scene and asked them to rate how informative they were (or

how well they could determine the scene content from this patch alone). They found that patches that coincided with places where people fixated were more semantically informative than random regions. Of course, if bottom-up saliency is correlated with "objectness" or semantic value then the two hypotheses, attentional guidance by top-down meaning or by bottom-up saliency, are hard to separate. As well as being fixated for longer, semantically incongruous objects may also be fixated earlier and this has been shown even when visual saliency is controlled (Underwood & Foulsham, 2006). Although this effect has not always been found it does seem that semantic factors encourage earlier inspection of objects and regions. On the other hand, there is a wealth of evidence suggesting that semantic analysis of words and other higher order tasks such as object recognition cannot be performed prior to fixation and so semantic variations should not have affected the time to reach the target. According to Henderson et al.'s (1999) saliency map hypothesis early eye movements are directed according to visual factors (such as bottom-up saliency) only, and semantic factors can only have an influence after the first fixation.

A closer inspection of the data can clarify some of these points. While reaction times were shorter for salient regions there was a large difference (1-2 seconds) between the time when the regions were first inspected and the manual response. This time was presumably occupied by processing the region and refixating it or elsewhere, and its length was probably confounded by the knowledge of the second task to locate where the region was. Could the difference in reaction times to different regions be explained by the amount of processing, encoding or localising which went on after the first fixation, rather than the time to find it? There are two reasons for thinking this is not the case. Firstly, there was also a difference of the same magnitude in the time to first fixate the object; people moved to salient regions earlier. Secondly, when looking at mean first fixation duration people spent longer processing salient regions than control regions, and as I have mentioned this may have been because they were more meaningful. Thus even though people spent longer fixating salient regions, they still responded quicker, and this must be mostly due to the speed at which the target was found.

It was not possible to fully separate the semantic and visual differences between the salient and control regions in these experiments. If, as seems likely, saliency is a visual short cut to the informative parts of a scene, then the two will coincide in many cases. If meaning cannot be processed in the periphery then the search benefit seen here must be due to visual saliency. It should be noted that if the peripheral processing of meaning underlies the saliency advantage this processing was equally uninterrupted by removal of high spatial frequency detail, contrast and colour. While information such as the gist of a scene can be acquired from low frequency information (Torralba et al., 2006) it seems unlikely that small objects and scene regions could be identified and influence attention top-down.

Another, knowledge-based explanation for the saliency advantage is that as these regions were probably more likely to contain objects participants may have known in advance where the target was likely to occur. This would also explain why the gaze contingent filtering did not change the benefit for salient regions. This would be an entirely top-down effect, and I tested this with two further experiments. Inverting the images might disrupt the way people use global layout, despite leaving saliency information unchanged. If the saliency effect is caused top-down then it might be reduced or removed in inverted scenes. There were signs that this was indeed the case—there was no reliable effect on the time taken to fixate the target, and this may have been due to the inversion manipulation. In target absent trials, the fifth most salient region was fixated more often than the first, a pattern that was not seen elsewhere. However, the RT was still faster for salient regions, which may have been due to processing time after the region had been fixated for the first time.

The final gaze contingent experiment offered a more extreme test of the effects of top-down expectations. If salient regions were still found quicker, even when there was no peripheral input to attract attention, then it would have to be due to knowledge about where that region is going to be. However, unlike all the other gaze contingent conditions in Experiment 5(a), salient regions were not found any quicker than control regions (in fact they were found slightly slower, which may have been because control regions were slightly closer to the centre on average). The effect of saliency on reaction

time was also removed. This is strong evidence that an account that explains the saliency advantage in terms of top-down expectations about where regions will occur cannot explain the data. Peripheral information is necessary for the effect to occur, which is consistent with a bottom-up account. It is slightly problematic that there was still an effect in target absent trials, where participants had not seen the regions before, and where they could not see them in the periphery. The fact that salient regions were still more likely to be fixated than controls is strange and suggests caution regarding this result in target absent trials elsewhere. This issue might be resolved by considering the area surrounding the different regions. Perhaps salient regions were more likely to be close to (or even part of) other important objects, and so when these parts were fixated and within the window, the gaze was more likely to move toward the region of interest. For example, a participant who happens to have found the bottom of a tree truck might then move up to fixate the branches, even though they were masked, and this might happen more often for salient regions.

Taken together, it seems that while top-down semantic factors have a role in eye guidance in this task, saliency does so also. The gaze contingent manipulations were useful in exploring these factors, and I will conclude this discussion by considering the impact of this type of display, and the implications for eye movement control in scenes.

7.5.3 The impact of gaze contingent filtering

Perhaps unsurprisingly, the gaze contingent display made search more difficult. This was reflected in lower accuracy, longer reaction times and longer times to fixate the target. The gaze-contingent window made search harder, and this is unsurprising given previous research showing more and longer fixations and shorter saccades (van Diepen & d'Ydewalle, 2003) in these types of display. Interestingly, the contrast-adjusted condition had little or no effect on search times relative to undistorted search. Globally lowering the contrast would preserve relative differences in intensity, and as it is these differences that are mostly important in the visual system perhaps this is not surprising. On the other hand, the blur and greyscale conditions selectively remove information (high spatial frequency and colour information respectively) so these have more impact in the efficiency of planning eye movements. The level of blur used was much higher than

contrast detection thresholds in the periphery, and if a less severe or variable blur were used, as in Loschky et al. (2005) then less of a search deficit would be predicted.

The effects on search efficiency illustrate the (perhaps trivial) point that eye movements are guided by peripheral vision. In Experiment 5(c), when there was no information present in the periphery, search took much longer because it had to proceed based only on the information at fixation and the participants' knowledge of scene layout, or else in a random, exhaustive fashion fixating the entire image. More interesting are the effects on fixation duration. What controls the length of time people fixate an object for? Following from Morrison's (1984) model, research in reading has assumed that a fixation is terminated once a criterion level of processing of the current word, for example lexical access, has been achieved. In models of scene perception it is less clear what makes up the fixation duration, and this is beyond the scope of a saliency map model. Henderson (1993) suggested that attention is disengaged once the object at fixation has been recognised and processed sufficiently, at which point a new saccade is programmed. If this were the case then the first fixation duration should be relatively unaffected by peripheral masking, particularly with a large window such as in the present experiments.

In fact, fixation durations were affected by peripheral filtering, and this replicates van Diepen and d'Ydewalle (2003). Presumably the distortion slowed selection of the next saccade target and thus delayed disengagement. This was particularly true in the greyscale condition, suggesting that colour is important in saccade target selection, despite not having an effect on region saliency. In some ways the removal of colour information is the least realistic manipulation—in the visual system contrast sensitivity and acuity decline with eccentricity, but the world does not change to black and white. It may be that in normal vision coarse, peripheral information can aid the identification of objects at fixation, so that for example an object in the context of a (congruent) scene is recognised more easily leading to a shorter fixation. When contextual information is removed recognition becomes harder, and so fixation is prolonged.

It is interesting to note that the completely masked conditions in Experiment 5(c) also led to a longer mean first fixation duration than the normal search, but that the

difference was less acute than when there was some information remaining (as in the blurred, greyscaled and adjusted-contrast conditions of Experiment 5(a)). This is surprising, as one would expect the remaining information to provide a benefit, rather than an impediment, relative to complete masking. This suggests that at least some of the difference in fixation durations is due to general task confusion and not what is being accomplished at fixation. Perhaps in a fully masked display participants tended towards more random eye movements, as they knew that they had no peripheral information to guide them. In Experiment 5(b) it was predicted that scene inversion might make objects harder to identify and so lead to longer fixations, but this was not the case. It may be that the demands at fixation in this task were relatively slight, and so fixation duration was more affected by information in the periphery than that at fixation.

The gaze contingent results were interesting, but further research with better controlled stimuli and tasks would be necessary to fully interpret the findings. With different degrees and combinations of filtering it would be possible to titrate exactly what information is necessary to plan and execute saccades efficiently. For the moment, the effects of peripheral filtering on fixation duration suggest that this measure is affected by the planning of subsequent saccades. This could be explained in the context of Findlay and Walker's (1999) model, in which there is a balance between the trigger to initiate a new saccade and the desire to maintain fixation. With more uncertainty about where to move to, the "when" fixate centre dominates over the "where" centre.

7.6 Conclusions and links forward

All three conditions of Experiment 5(a) showed an advantage for saliency and thus provide partial support for a saliency map model of eye movements in natural scenes. This advantage occurred in a search task and so was not completely overridden by top-down instructions. However, there were some aspects of the findings that a bottom-up model would find difficult. Medium saliency regions fixated after five predicted fixations showed no consistent above-chance advantage suggesting that after initial fixations the model has poor predictive power. Gaze-contingent peripheral filtering did not ameliorate the saliency effect, and this is problematic for a model that depends on bottom-up, peripheral information controlling attention. The picture becomes clearer when top-down control and semantic influences are considered. Future research will need to separate the effects of bottom-up capture, top-down guidance and processing at fixation in order to provide a better account of where people look. This is hard in natural scenes, especially if saliency and informativeness naturally co-occur. In the next two chapters I use more controlled stimuli and return to the influence of saliency on search.

8 Saliency and set size in an object search task

8.1 Introduction

Previous experiments in this thesis have explored the effect of target saliency on visual search for both a target object (Experiment 1) and a region within a photograph (Experiments 4 and 5). This chapter describes two experiments using a more controlled stimulus—an array of photographs of objects. The experiments ask whether the saliency map model is a useful predictor of how quickly an object will be found.

Visual search is one of the most studied tasks in experimental psychology (see Wolfe, 1998a, for a review). Typically, the procedure is as follows. Participants are told about or shown a target item that they then have to search for in an array of similar items. The task is to confirm its presence or absence, amid several non-targets or distractors. Models of this paradigm (in particular FIT, Triesman & Gelade, 1980) have focused on explaining search efficiency in terms of the featural similarity between target and distractors and the number of items in the display—the set size. "Feature search", where the target is identified by a single feature, results in an easy, highly efficient search which is relatively unaffected by the set size. In other words, the slope relating set size to the time required to find the target is flat and the target "pops out" regardless of the number of distractors. In a more difficult, "conjunction search" where the target is defined by the conjunction of two or more features (such as finding a red 'T' amongst green 'T's and red 'L's) the search time increases with the number of distractors, suggesting that each item must be inspected serially.

In Itti and Koch (2000) the authors introduce the saliency map model as a model of visual search. As well as a more natural search task looking for vehicles in landscapes, they suggest that their model can account for the search slopes seen in feature and conjunction search. The model therefore makes predictions about how search proceeds in both complex natural scenes and more simple artificial displays. Part of the discrepancy in the results of previous experiments might be due to the difficulty in controlling aspects of natural photographs. For example, some scenes might contain more clutter—more items to search—than others. It has also been shown that the context of a realistic scene

provides top-down information, such as coarse information about layout and where real objects are likely to appear, which can guide search (Neider & Zelinsky, 2006). When these constraints are violated, search is disrupted. In previous chapters I have suggested that these constraints can override the effects of low-level saliency. It is thus useful to test the model in a more controlled stimulus and to see whether the effects of saliency are clearer.

What will the effect of saliency be on a visual search task? The (purely bottom-up) saliency model predicts the order in which items will be fixated, and so more salient targets should be fixated earlier, and lead to faster response times. There is evidence that highly salient singletons capture attention in simple visual search displays (Nothdurft, 2002; van Zoest & Donk, 2005). However, when top-down information about the target was available in Experiment 1(a), target saliency did not make much of a difference. As discussed previously, and as suggested by models such as those by Rao et al. (2002) and Navalpakkam and Itti (2005), knowledge of target features can bias search. In this chapter the saliency of the target will be manipulated. If saliency is not important in a search task then salient and non-salient targets will be found on the basis of their features and just as easily.

Two further questions are asked. First, does target saliency change the search slope relating search and set size? Targets in these experiments are relatively complex objects defined by the conjunction of several features. If saliency has an effect on covert attention and on the guidance of fixations, then it might lead to a less steep search slope because the target will capture attention and items will not have to be searched serially. A non-salient target should not pop out and so a larger set size will provide more items to search before inspecting the target. Previous experiments looking at search in natural scenes have found it difficult to quantify the set size in a continuous scene that cannot be easily split into objects of interest. Some recent research addresses this by developing measures of clutter or set size in scenes (Neider & Zelinsky, 2008). The object arrays used here are much easier to divide into search items and so are useful for exploring this problem.

A second point of interest is whether the amount of information about the target influences search or the effect of saliency. In Experiment 1a, it was hypothesised that searching for a specific exemplar might be more efficient than searching for a general category. This would also provide more top-down information to bias a saliency map representation of the features in an image. However, there were no differences. In this chapter, the first experiment gives people a verbal description of the search target, while in the second a pictorial cue is given. Search is faster when a more exact target cue is provided (such as an accurate picture rather than a semantic label), and this is true with both simple line stimuli (Wolfe et al., 2004) and complex objects (Vickery et al., 2005). However, it is not known how cue type and saliency interact to affect eye movements. A pictorial target template should provide more top-down guidance and so it might reduce the effect of bottom-up information.

8.2 Experiment 6: searching for a verbal target

8.2.1 Method

Participants

Eighteen student volunteers took part. All reported normal or corrected-to-normal vision and passed a calibration and validation procedure.

Stimuli, apparatus and design

A set of stimuli was created which would be relatively simple in terms of the distinct items present, but which contained realistic objects in a plausible setting. To do this a variety of items of food and drink were photographed front-on against a plain white background. This gave a set of objects which were of a similar resolution and which were lit in a similar way. A set of 30 objects was compiled, all of which were of approximately the same size and which were easily identifiable. The objects were extracted from their background and resized to fill a standard sized square, which in this experiment had dimensions of 4.5˚. A short verbal label was paired with each object, to be used as a target cue (for example, “a bottle of milk”).

The search arrays were composed of several objects positioned on a set of shelves. This backdrop was chosen to provide a more naturalistic framework. The

shelves were simplified from a real photograph and had four horizontal levels. The whole image subtended 34˚ by 27˚. Objects were pasted into set positions using Adobe Photoshop 7.0. There were 18 possible positions, based on a grid, and the horizontal distance (between the centres of adjacent object positions) and the vertical distance between levels subtended approximately 6˚.

Objects were placed randomly in these positions to create many possible stimuli. No specific object appeared more than once in the same array. Three set sizes were used, with 8, 12 and 16 objects in each. Figure 8.1 shows several examples. To make the arrays less homogenous, the colour of the background was varied. This meant that some objects became more salient than others, due to the changing colour contrast between objects and background. In order to quantify the saliency of objects in the arrays, the complete images were processed by the saliency map model, resulting in a series of model-predicted fixations (see Figure 8.3 for an example). This model output was used to select the final stimuli, and the target in each case, based on the following design.

There were 120 experimental images, half of which were designated target-present stimuli. These stimuli were divided equally into the three set sizes and according to whether the target object in each case was salient or non-salient. Salient targets were defined as those that were selected on the first shift of attention by the saliency map model, and thus these were the most salient objects in the scene. Non-salient targets did not get selected within the first five model fixations, showing that they were not among the most conspicuous objects. The shelves and array background were designed to be not highly salient, and indeed it was very rare for the model to select anything but one of the objects. The target object, and the position which it occupied in target present trials, was chosen at random from the full set. In order to ensure that it was not more efficient to respond to certain objects or positions, every object was equally likely to be present or absent when it was the target. In addition, the target occurred in each position about the same number of times (2 to 3 times in each position).

All participants saw the same set of stimuli in a randomised order. The two factors of set size (8,12 and 16 objects) and target saliency (salient and non-salient) were

thus manipulated within-subjects. The EyeLink II tracker was used with the set-up described previously.

Figure 8.1. Some examples of the stimuli from Experiments 6 and 7. A variety of realistic objects appeared at randomised locations on a set of shelves and a coloured background. Three set sizes were used: 8 items (top), 12 items (middle) and 16 items (bottom).

Procedure

The procedure began with a calibration and written instructions, followed by a short practice session of 16 trials, which followed the same sequence as the experimental trials but contained pictures that were not used again. In each trial (Figure 8.2), the verbal cue was presented as black text in the centre of a plain white screen. The text was replaced after a fixed duration of 2000 ms with a drift correct marker. Previous experiments had observed effects of the starting location in search experiments, so it was decided to vary the location of the drift correct marker. Six locations were used, aligned on the top and bottom rows of an evenly spaced, imaginary 3 x 3 grid. The location of the drift correct was determined pseudo-randomly, with the constraint that it never appeared in exactly the same position as the target in the following trial. The experimenter confirmed when a stable fixation was maintained on this location, and the search array was presented. Participants were instructed to press a key to indicate the presence or absence of an object that matched the verbal description, as quickly as possible. Some subjects were unsure of the identity of all objects, but they were instructed to make their best guess. The search array was offset by the response, and the next trial began. Testing proceeded in three blocks of 40 trials, and participants were recalibrated and given a short break in between each block.

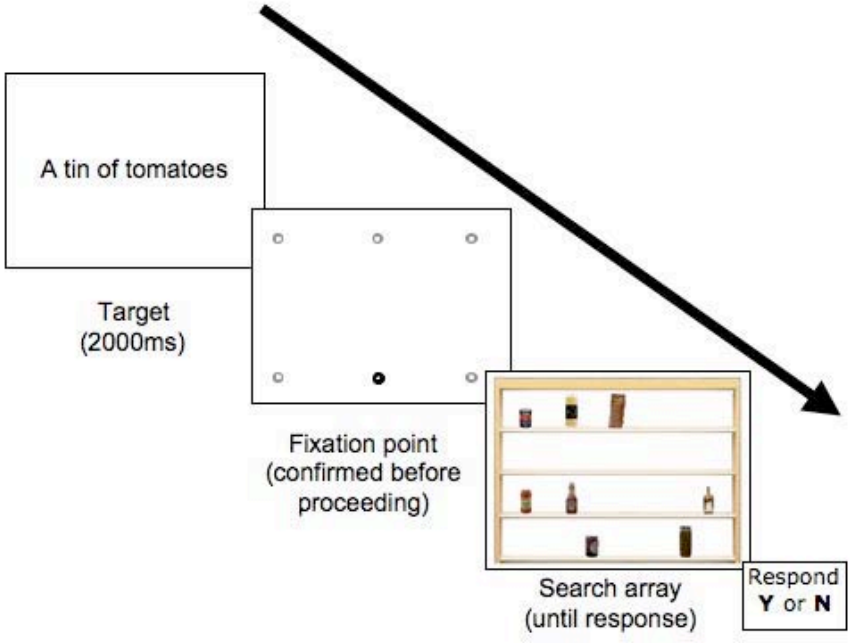


Figure 8.2. The procedure for each trial in Experiment 6. The target and the location of the fixation point varied from trial to trial.

8.2.2 Results

Response measures

The manual responses were first inspected to determine whether set size and target saliency affected the accuracy or speed of search. Error rates were low overall, with incorrect responses occurring on only 9% of trials, on average, across all trials. The proportion of correct responses and the reaction time (RT, for correct trials only) are shown in Table 8.1 for the different target present conditions, and for target absent trials.

The low error rate, particularly in target absent trials, suggests that participants were able to identify objects based on the label: their interpretation matched that of the design. In terms of accuracy, there were few differences between the conditions. Reaction time was prolonged in target absent trials, and in general it increased with larger set sizes. Target absent trials were responded to slightly more accurately (people were reluctant to give false alarms) and more slowly than target present trials (a common finding in visual search, Wolfe, 1998a). As the focus of this experiment is the effect of target saliency, analyses will look only at the target present trials.

Target type		Set size 8 Non-salient	8 Salient	8 TA	12 Non-salient	12 Salient	12 TA	16 Non-salient	16 Salient	16 TA
Proportion correct	***M***	0.86	0.92	0.95	0.91	0.91	0.94	0.88	0.85	0.97
	SEM	*0.02*	*0.02*	*0.02*	*0.02*	*0.02*	*0.02*	*0.03*	*0.03*	*0.01*
RT (ms)	***M***	1440	1509	1988	1490	1456	2156	1867	1664	2569
	SEM	*78*	*56*	*54*	*72*	*64*	*65*	*97*	*74*	*81*

Table 8.1. Response measures from Experiment 6.

The results were analysed with 2 (levels of target saliency) x 3 (set sizes) repeated-measures ANOVA. There were no reliable main effects on accuracy: main effect of target saliency, $F(1,17)<1$; set size, $F(2,34) = 1.74$, $MSE = 0.0078$, $p = .19$. There was a marginal interaction, $F(2,34) = 3.11$, $MSE = 0.0066$, $p = .057$, which came about because salient targets were responded to more accurately than non-salient targets in the smallest set size, but less accurately when there were 16 objects in the array. Looking at reaction time, there was a large effect of set size, $F(2,34) = 14.74$, $MSE = 69524$, $p<.001$, but no reliable difference between targets of different saliency, $F(1,17) = 2.35$, $MSE = 36207$, $p = .14$. Responses were slower to arrays containing 16 objects than to the smaller set sizes, both ts(17)>4, ps<.001, although these did not differ. There was a reliable interaction, suggesting that the RT slope associated with increasing set size was different for the two types of target, $F(1.4,24.4) = 4.04$, $MSE = 58390$, $p = .042$, with Greenhouse Geisser correction. Reaction time increased more steeply with set size when the target was non-salient: simple main effect of set size, $F(2,34) = 14.13$, $MSE = 69524$, $p<.001$; than when it was salient, $F(2,34) = 3.03$, $MSE = 69524$, not reliable at $p = .061$.

Eye movement measures

In general, participants made 6 to 10 fixations in target present trials, and the mean duration of their fixations was similar to that reported elsewhere in this thesis.

Of most relevance for the task at hand is the speed at which participants located the target in each condition. Analysis began by looking at the first fixation time on the target, which was defined as the time from the onset of the picture to the start of the first fixation within the target region. This analysis excluded trials which were incorrect, and those where the target was not fixated at all (these were rare and normally resulted in an incorrect response: the target was fixated in 93% of correct, target present trials). The first fixation time is plotted in Figure 8.3, and there is a clear search slope, with targets taking longer to find with a larger set size. The effect of set size was reliable, $F(2,34) = 27.1$, $MSE = 26866$, $p<.001$, but there was no effect of target saliency, $F(1,17)<1$, and no interaction, $F(2,34) = 1.4$, $MSE = 47584$, $p = .26$. All pairwise comparisons between the marginal means for each set size were reliable (all ts$(17)>4$, ps$<.005$). It took approximately 700 ms for participants to first fixate the target when the array contained 8 objects, and the time increased steadily with the two larger set sizes at a rate of about 36ms per extra item in the display.

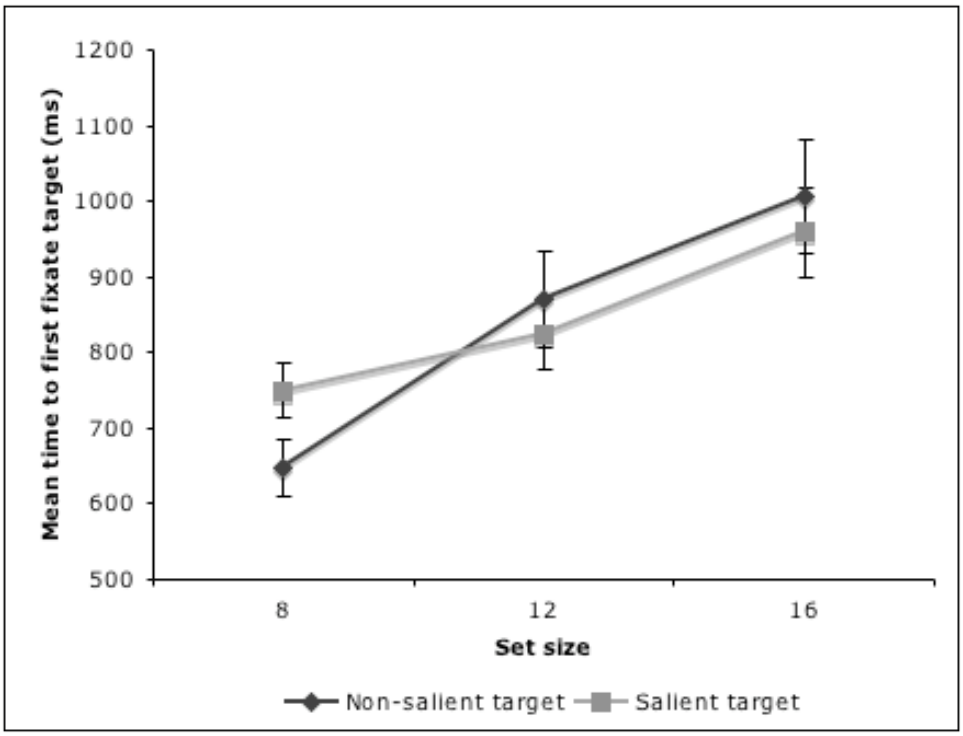


Figure 8.3. The mean time taken to first fixate the target, as a function of the set size and the saliency of the target. Error bars show the standard error of the mean.

The pattern of results in first fixation time was somewhat different to that in reaction time, and there is a surprisingly large difference between the two measures. This suggests that participants made several more fixations or spent some time inspecting the target before making their response. It is important, therefore, to look at the time taken to process the object. To rule out the possibility that salient targets were easier to identify or respond to once fixated, I analysed the first fixation duration on the target. There was no effect of set size, $F(2,34)<1$; or of saliency, $F(1,17) = 2.2$, $MSE = 1508$, $p = .15$; and no interaction, $F(2,34) = 1.82$, $MSE = 2430$, $p = .17$. Most important, the difference between the mean first fixation duration on salient (280ms, $SEM = 10.1$) and non-salient targets (269, $SEM = 12.7$) was negligible.

In Experiment 1(b), the first saccade towards a target came from further away when that target was salient. This suggested that salient objects could attract covert attention and trigger an eye movement from further away more than non-salient objects. To address this in the current experiment the first saccade to land on the target was inspected. This did not include any trials where the target was not fixated, and it looked at only the first saccade to enter the target area in any one trial. The amplitude of this saccade was reliably greater for salient ($M = 9.4^{\circ}$, $SEM = 0.46$) than non-salient targets ($M = 7.8^{\circ}$, $SEM = 0.34$), $F(1,17) = 7.89$, $MSE = 8.25$, $p = .012$. There was no effect of set size on this measure, and no interaction: both Fs$(2,34)<1$. A possible problem with interpreting this effect is that, because the location of the fixation point where search started was varied, the difference in saccade lengths could have come about from a systematic difference in the distance between start point and target across conditions. This would not be a straightforward explanation because normally several saccades were made within a trial so starting location was unlikely to have a large effect. Nevertheless, to check this possibility I compared the Euclidian distance between the fixation point and the target location for salient and non-salient targets. There was no reliable difference between the two types of target (salient, $M = 16.2^{\circ}$, $SEM = 1.1$; non-salient, $M = 14.8^{\circ}$, $SEM = 1.3$; independent-samples t-test, $t(58)<1$).

Scanpath analyses

In Chapters 4 and 5 some techniques for quantifying the sequence of fixations on a scene was discussed. The search arrays in the current experiment were composed of several distinct objects, so it is interesting to ask what determined the order in which these objects were inspected. As outlined in the method, the saliency map model made predictions about the order in which objects should be selected. How closely did the experimental scanpaths observed resemble the model-generated sequence?

To answer this question, each object was labelled with a character (starting with "A" for the leftmost object on the top shelf) and the saliency ranking of the first five objects in the scene was transformed into a letter string (see Figure 8.4). The first five model shifts were used because at least this many were available for all the stimuli, objects were not re-selected within this number (due to the inhibition of return in the model), and participants also tended to fixate at least this number of items. For the observed data, each fixation that was on an object was labelled with the corresponding letter. This did not include the very first fixation, which started before the onset of the search array and which was dependant on the drift correct location. Fixations which were not located in an object region were very rare and so not included, and multiple, consecutive fixations on the same object were condensed into one. The participant scanpaths were trimmed to a length of 5 items. The scanpath and saliency strings were then available for comparison using the string-edit method which gives an index between 0 and 1 for completely dissimilar and identical strings respectively (see Chapters 4 and 5 for full details).

Three sets of comparisons were performed. First, the similarity between the participant scanpaths and the saliency ranking for each trial was assessed. This gives a measure of how well the sequence of inspection is predicted by the model. There was no reason to believe that saliency was systematically distributed in the search arrays and so if this scanpath comparison is more similar than sequences drawn at chance then it would suggest a link between saliency and the search sequence. To assess whether participants were systematic in the order in which they searched the objects two further comparisons calculated the similarity between scanpaths recorded in the experiment.

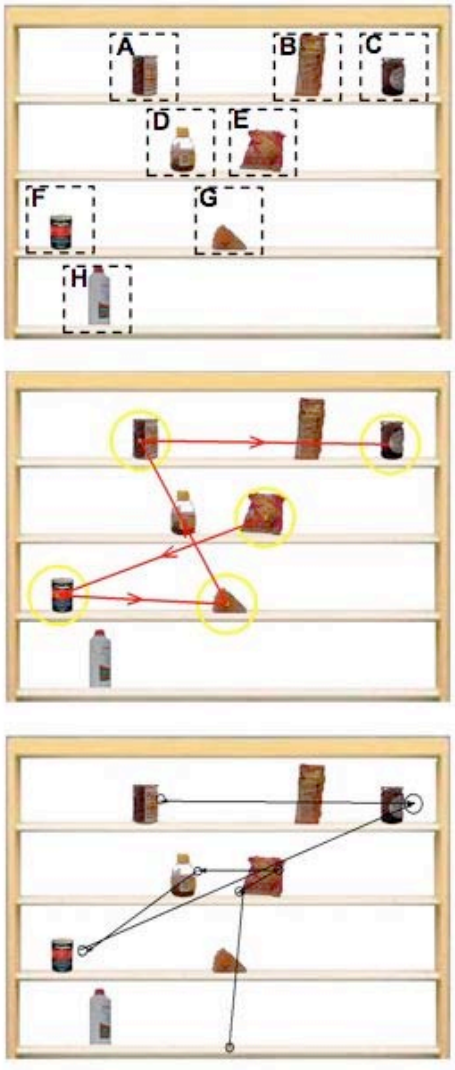


Figure 8.4. Scanpath analysis for one stimulus in Experiments 6 and 7. Each object was labelled with a letter, starting at the top left of the array (top panel). The scanpath from the saliency model (middle) and from an observer (bottom) were transformed into character strings ("EFGAC" and "EDFCA" in these examples). These strings have a string-edit similarity of 0.4.

The "within-subject" similarity was calculated as the mean similarity between a scanpath from one trial and those from all other trials made by the same person viewing stimuli from the same set size. The "within-trial" similarity was the mean similarity between scanpaths from different participants viewing the same array. All three sets of comparisons were performed separately for each condition, and also for target absent trials. The number of possible scanning sequences, and therefore the similarity expected by chance, varies with the set size. For example, with 12 objects there are more possible sequences in which any 5 items can be inspected than with 8 objects, and so the similarity

to be expected by chance will be lower. As explained in Chapter 4, this can be estimated by generating random strings. To compute the string-edit similarity expected by chance 10000 pairs of strings were randomly generated with the correct number of possible objects, and these were compared. The mean chance expectancy was 0.150, 0.097 and 0.070 for a set size of 8,12, and 16 items respectively. Note that these values are close to what could be computed mathematically, $1/n$ where n is the number of items (they are different from this due to the constraint that two consecutive fixations could not be on the same object, which reduces the number of potential sequences).

In order to compare between different set sizes it was desirable to take the chance expectancy into account. This was accomplished by computing a chance-adjusted similarity, which was simply the difference between the mean similarity score and the chance value for that set size. If this value is positive then the scanpaths being compared are more similar than expected by chance. Figure 8.5 shows the chance-adjusted similarity according to this procedure, for each of the three comparisons and across the different conditions. Several trends are evident from this figure. Firstly, it is clear that the similarity is greatest in the within-trial comparison and it is only in this case that all conditions are positive (i.e. above chance). Looking at the comparison between the observed data and the saliency predictions, it is mostly when the target is salient that the similarity is consistently greater than chance. In the within-subject comparisons, the similarity between different scanpaths made by the same person is only above chance at the largest set size.

The data were analysed in the following two ways. First, repeated-measures ANOVA looked for a difference between conditions (3 types of trial x 3 set sizes). Then, in order to see which comparisons were greater than chance (that is reliably positive) I performed a series of one-tailed, one-sampled *t*-tests that tested the hypothesis that the chance adjusted similarity was greater than zero. Given that this involved a large number of *t* tests (9 per comparison), I used a conservative, Bonferonni correction for family-wise error. As such I will only report those *t* tests that gave a *p* value of less than 0.0056 (these conditions are marked in Figure 8.5).

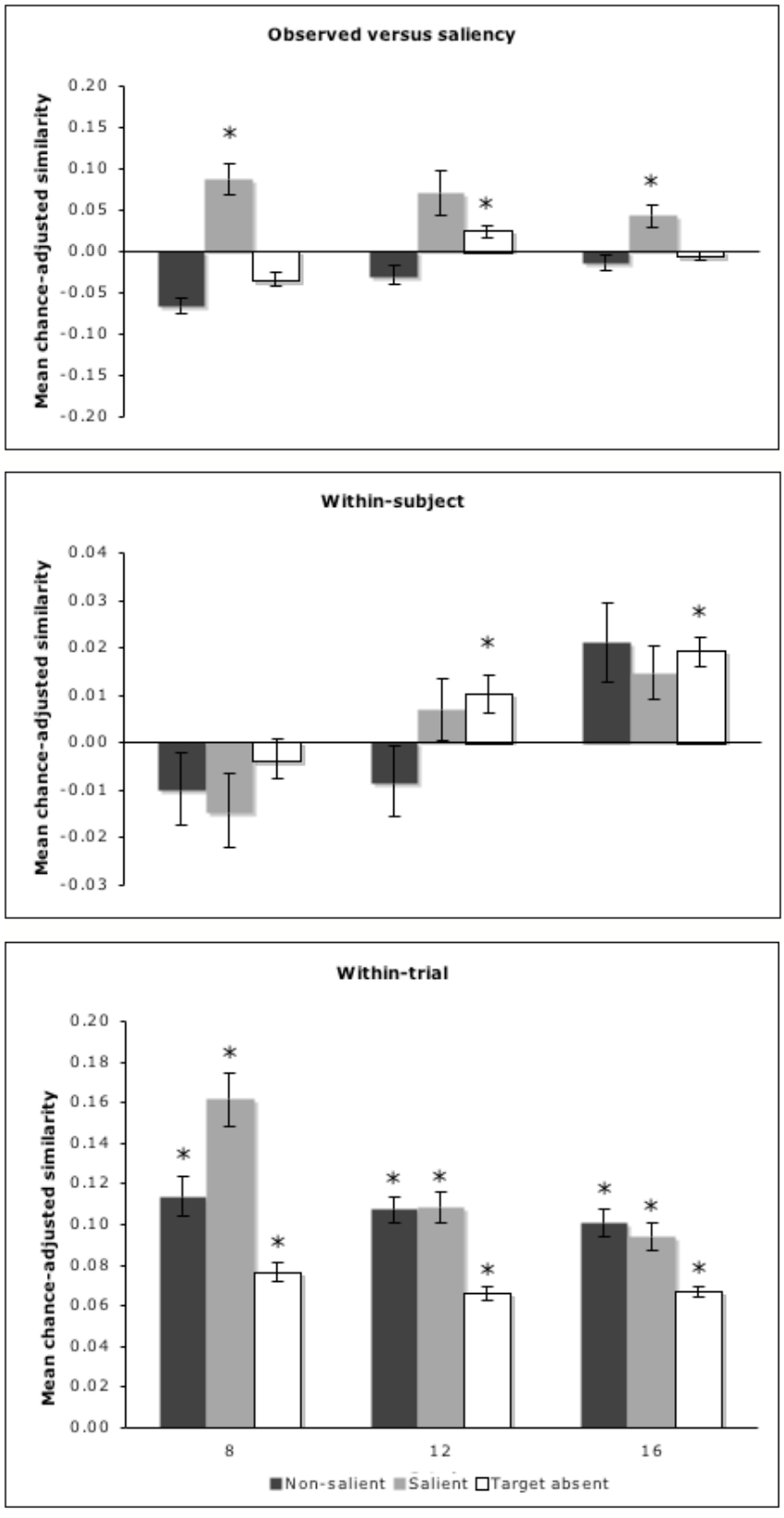


Figure 8.5. The mean and standard error, chance-adjusted similarity between scanpaths. The comparisons were between participants' scanpaths and the model predictions (top), scanpaths from the same person looking at different stimuli (middle) and scanpaths from different people looking at the same array (bottom). The similarity score plotted is the difference between the mean similarity and the chance value for that set size. Asterisks indicate values reliably greater than zero (see text).

There was no reliable effect of set size on the similarity between the model and observed scanpaths, $F(2,34 = 2.80$, $MSE = 0.0031$, $p = .075$ (top panel, Figure 8.5). When chance was taken into account, it was not the case that the model was better or worse at predicting scanpaths with different numbers of items. There was, however, a large effect of trial type, $F(2,34) = 36.47$, $MSE = 0.0040$, $p<.001$, and a reliable interaction, $F(4,68) = 4.13$, $MSE = 0.0034$, $p = .005$. Comparisons between the different types of trial indicated that all differences were significant: salient target-present trials resulted in greater similarity than both those with non-salient targets, and those where the target was absent. Non-salient target-present trials led to the lowest similarity (all comparisons $t(17) > 3.4$, $p < .01$). One can see from the graph that the interaction was due to similarity increasing with set size in trials with a non-salient target, but not elsewhere. Specifically, there was a simple main effect of set size in non-salient target present trials, $F(2,34)=4.2$, $MSE=0.003$, $p<.05$, due to a reliable difference between stimuli with 8 items and the larger set sizes, both ts$(17) > 2.7$, ps $< .05$. There was no reliable effect in salient target trials, though there was an effect in target absent trials, $F(2,34)=4.8$, $MSE=0.003$, $p<.05$, where 8-item arrays produced less similar scanpaths than 12-item arrays, with 16-item arrays intermediate between the two, all ts$(17) > 2.8$, ps $< .05$. Of all the conditions, the similarity was significantly greater than chance in only 3 cases: salient target-present trials with 8 items and with 16 items, and target-absent trials with 12 items ($t(17) = 4.52$, 3.20 and 3.02 respectively). Thus, both the ANOVA and the comparisons against chance showed that the model better predicted scanpaths made in trials with a salient target. However, this is not surprising as in these trials the task demanded that participants inspect at least one salient item (the target), which was by definition predicted by the model. By the same token, non-salient trials provided a top-down incentive to look at a non-salient region, which would explain why the saliency scanpath was not similar to that made by participants. The strongest support for the model would be similarity between the model-predicted scanning sequence and the human-generated scanpath in target-absent trials, where, perhaps, the top-down signal was weakest. However, the similarity in this case was low. In other words, the model was not a reliable predictor.

Looking at the within-subject comparisons (Figure 8.5, middle panel), there was a reliable effect of set size, $F(2,34) = 12.97$, $MSE = 0.0008$, $p<.001$, but no main effect of trial type $F(2,34) = 1.12$, $MSE = 0.0007$, $p = .34$, and no interaction, $F(4,68) = 1.45$, $MSE = 0.0006$, $p = .23$. The largest set size resulted in a greater within-subject similarity than stimuli with either 8 or 12 items, $t(17) = 5.11$ and 3.48 respectively, both ps<.01. Consistent with this finding, the similarity was not greater than chance in any of the 8-item conditions but it was above chance in stimuli with a set size of 16. However, the differences were small and they were only reliable in the target-absent trials with 12 and 16 items, $t(17) = 2.94$ and 5.81 respectively. These comparisons are based on the scanpaths made by individuals viewing different stimuli and thus idiosyncratic, systematic scanning patterns did not appear to play much of a role in the sequence in which objects were fixated, with the possible exception of the largest set size.

The similarity within trials was much greater (Figure 8.5, bottom panel), showing that different people viewed the same stimulus in a way that was more similar than expected by chance. All one-sample t-tests were highly reliable: all ts(17)>11.5, ps<.001. There were reliable main effects of both set size, $F(2,34) = 11.74$, $MSE = 0.0011$, $p<.001$, and trial type, $F(2,34) = 54.45$, $MSE = 0.0007$, $p<.001$. Pairwise comparisons showed that scanpaths were more similar in the 8-item stimuli than elsewhere: both ts(17)>3, ps<.05, whilst the two larger set-sizes did not differ. All three different types of trial were reliably different, all ts>2.6, ps<.05. Mean similarity was greatest in target-present trials with salient targets, followed by those with non-salient targets. Target absent trials resulted in the lowest similarity between participants viewing the same stimuli. The main effects were qualified by an interaction, $F(4,68) = 6.84$, $MSE = 0.0008$, $p<.001$. A larger set size reduced the similarity within salient, target-present trials, but not in other trial types (simple main effect of set size in salient target trials, $F(2,34)=20.1$, $MSE=0.023$, $p<.001$; not reliable in non-salient and target trials, both $F(2,34)<1$). How should the within-trial comparisons be interpreted? They reveal that there were patterns in the objects fixated in any one trial. This is likely to be due to top-down factors: the similarity was greater in conditions where there was a target, presumably because people consistently fixated this object. That there was significant

similarity in the target absent case indicates that people did not just randomly scan the objects but were following a pattern. However, as I have shown, this was not well described by saliency.

8.2.3 Conclusions

The main results of this experiment were: (1) that manual responses were slower to displays containing a greater number of objects; (2) that this slope was not reliable when the target was highly salient; (3) that the time required to fixate the target was affected by increasing the number of items; (4) that saccades to the target came from further away when that target was salient; and (5), that scanpaths were similar between individuals viewing the same array, even when the target was not present, but that this similarity could not be accounted for by saliency. I discuss these findings fully at the end of this chapter. Model-predicted saliency did have a small effect on reaction time, in combination with set size, and on saccadic amplitude, but otherwise salient targets were not necessarily easier to find than non-salient ones.

Experiment 7 replicated the same method as used here, but provided a picture of the target, in addition to a verbal description, to investigate the effect of a more precise target template on the search process. I predicted that search would be more efficient, and that this might lead to less of an influence of saliency.

8.3 Experiment 7: searching for a pictorial target

8.3.1 Method

Participants

Sixteen volunteers took part in this experiment. None of the volunteers had taken part in Experiment 6, and all had normal or corrected-to-normal vision.

Stimuli, apparatus and design

The EyeLink 2 eyetracker was used for this experiment, and the stimuli and design were exactly the same as in Experiment 6. In addition to the verbal description, the target screen in this experiment showed the object mounted against a white background and in the centre of the screen.

Procedure

The experiment proceeded in exactly the same way as the previous experiment. Following calibration and a practice session, participants saw all three blocks of 40 trials each, in a fully random order. Each trial consisted of a target screen, showing both a picture and a written description of the target (see Figure 8.6). This was followed by a drift correct marker in one of the six possible starting locations, and then a search array.

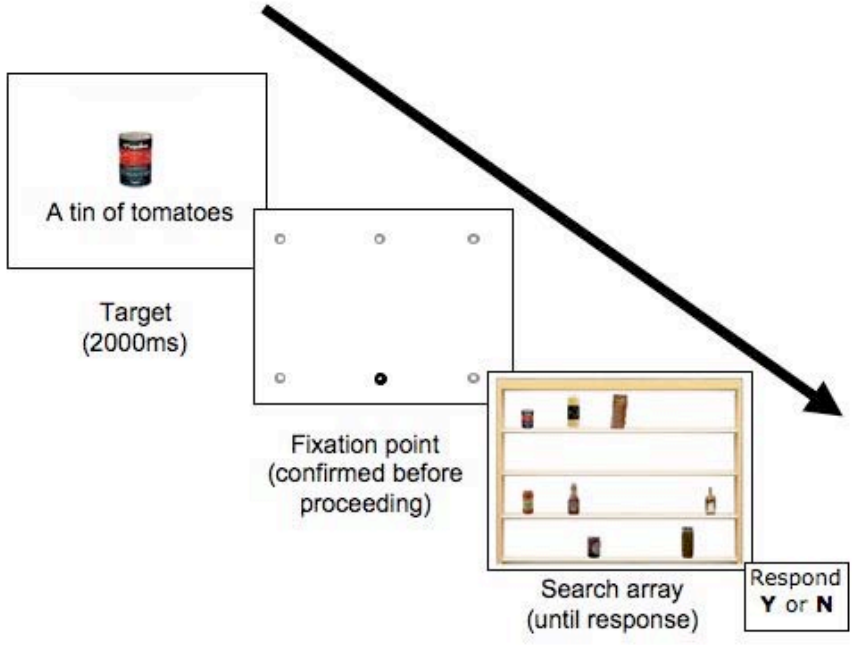


Figure 8.6. The procedure for one trial in Experiment 7.

8.3.2 Results

The same measures were taken as in the previous experiment, and these are summarised in Table 8.2. Statistical analysis used repeated-measures ANOVA. A few key measures were compared with the previous experiment using between-groups *t* tests.

Response measures

Looking first at the manual search responses, accuracy was again very high, with few misses and even fewer false alarms. In target present trials, there were no effects of target saliency $F(1,15) = 2.9$, $MSE = 0.0035$, $p = .11$, or set size, $F(2,30) = 2.8$, $MSE = 0.0038$, $p = .077$, on the proportion of correct responses, and no interaction, $F(2,30)<1$. Reaction times showed a similar pattern to those in Experiment 6: target absent trials were responded to more slowly and set size had a reliable effect in target present trials, $F(2,30) = 5.7$, $MSE = 20278$, $p<.01$. Whilst the RT to trials with 8 and 12 objects did not differ reliably ($t(15)<1$), stimuli with a set size of 16 led to a significantly longer RT (versus 8, $t(15) = 2.65$, $p<.05$; versus 12, $t(15) = 3.07$, $p<.01$). Saliency did not have a reliable effect on reaction time, $F(1,15)<1$, and there was no interaction, $F(2,30) = 1.41$, $MSE = 22869$ $p = .26$. In order to see whether the change in procedure in this experiment affected performance, a between-groups t test compared the mean reaction times in Experiments 6 and 7, collapsed across target present trials. RTs were significantly faster in this experiment ($M = 1421$ms, $SEM = 76$) than in Experiment 6 ($M = 1793$ms, $SEM = 54$, $t(32) = 4.04$, $p<.001$), indicating that giving a pictorial target made the search task easier.

		Set size								
		8			12			16		
Target type		**Non-salient**	**Salient**	**TA**	**Non-salient**	**Salient**	**TA**	**Non-salient**	**Salient**	**TA**
Proportion correct	***M***	0.96	0.94	0.98	0.93	0.91	0.98	0.96	0.94	0.97
	SEM	*0.01*	*0.02*	*0.01*	*0.02*	*0.02*	*0.01*	*0.02*	*0.02*	*0.01*
RT (ms)	***M***	1167	1238	1645	1184	1213	1676	1331	1278	2061
	SEM	*79*	*68*	*91*	*63*	*74*	*103*	*68*	*77*	*130*
Time to first fixate target (ms)	***M***	554	589		626	664		726	688	
	SEM	*26*	*25*		*24*	*38*		*29*	*37*	
Duration of first target fixation (ms)	***M***	256	255		283	272		270	278	
	SEM	*11.6*	*13.2*		*15.4*	*13.8*		*18.4*	*12.8*	
Mean amplitude of first saccade to target (°)	***M***	7.21	9.02		7.87	9.36		8	8.74	
	SEM	*0.3*	*0.5*		*0.41*	*0.44*		*0.49*	*0.53*	

Table 8.2. Summary of the main measures taken in Experiment 7.

Eye movement measures

The time to first fixate the target was lower in this experiment than in Experiment 6: it took a mean of just 641ms (*SEM* = 38) to make a fixation in the target region, compared to 843ms on average (*SEM* = 22) in the previous experiment. This is consistent with the reaction time data, and the between-groups effect was reliable ($t(32) = 4.46$, $p<.001$). In this experiment, time to target was not affected by saliency, $F(1,15) < 1$, but there was an effect of set size, $F(2,30) = 19.5$, $MSE = 7591$, $p<.001$. Arrays with 8 objects were responded to reliably quicker than those with 12 objects ($t(15) = 3.37$, $p<.005$) and those with 16 objects ($t(15) = 7.38$, $p<.001$). Trials with a set size of 16 had the longest time to target, and this was significantly different to the 12-object arrays ($t(15) = 2.53$, $p = .023$). The increase in time to target with set size was linear, but the slope was less steep than that seen in Experiment 6, at around 17s per additional item. There was no interaction between saliency and set size, $F(2,30) = 1.71$, $MSE = 8598$, $p = .20$.

As in the previous experiment, the mean first fixation duration on the target did not vary amongst the experimental conditions (no effect of set size, $F(2,30) = 2.22$, $MSE = 1987$, $p = .13$; saliency, $F(1,15)<1$; interaction, $F(2,30)<1$). However there was again an effect of saliency on the size of the first saccade to land on the target, $F(1,15) = 16.2$, $MSE = 2.67$, $p = .001$. There was no effect of set size on the amplitude of the first target saccade $F(2, 30)<1$, and no interaction, $F(2,30) = 1.13$, $MSE = 2.12$, $p = .34$. Saccades to salient targets were reliably longer than those to non-salient targets.

Scanpath analyses

The scanpath comparison method and the comparisons performed were exactly the same as in the previous experiment. Figure 8.7 plots the chance-adjusted similarity values in each case, and the pattern is extremely similar to that from Experiment 6.

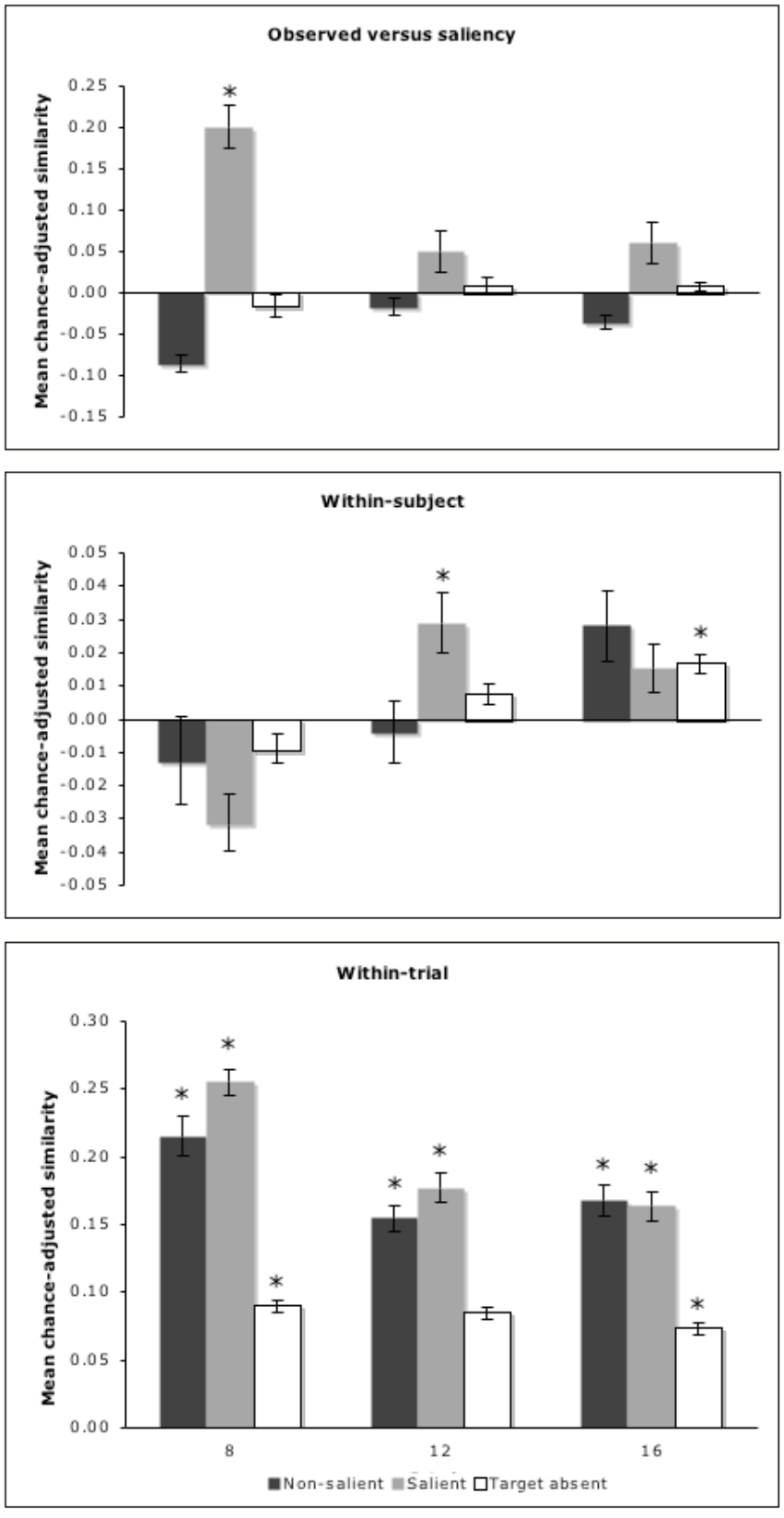


Figure 8.7. The mean, chance-adjusted similarity between scanpaths in Experiment 7. See legend from previous figure for more details.

Statistical analysis was performed as previously. Looking first at the similarity between the participant's scanpath and that generated by the model (Figure 8.7, top), there was no effect of set size, $F(2,30) = 1.58$, $MSE = 0.004$, $p = .22$, but a reliable effect of trial type, $F(2,30) = 66.97$, $MSE = 0.004$, $p<.0001$. Trials with a salient target led to higher similarity than both those with a non-salient target and those where the target was absent (both ts(15)>8, ps<.001). Scanpaths from target-absent trials were reliably more similar to the saliency scanpath than those with a non-salient target ($t(15)=5.3$, $p<.001$). This effect was qualified with an interaction between set size and saliency, $F(4,60) = 13.25$, $MSE = 0.005$, $p <.001$). For salient target trials, a larger set size led to lower mean similarity scores (simple main effect, $F(2,30)=28.1$, MSE=0.004, $p<.001$). However, for non-salient and target absent trials, similarity was higher with more items (simple main effects, $F(2,30)=5.2$, $MSE=0.004$, $p<.05$ and $F(2,30)<1$ for non-salient and target absent trials respectively). In terms of the absolute value of the similarity scores, only the salient target trials with 8 items led to a similarity estimate which was reliably greater than chance (with adjusted $p<.0056$; $t(15) = 7.61$).

The within-subject comparisons (Figure 8.7, middle) gave a reliable effect of set size, $F(2,30) = 15.03$, $MSE = 0.001$, $p<.001$, but no effect of trial type, $F(2,30)<1$. The scanpaths made by the same person viewing different arrays were less similar (and below chance) when there were only 8 objects, than when there were 12 or 16 (both ts(15)>4.8, ps<.005). The two larger set sizes did not differ reliably. This effect interacted with trial type, $F(4,60) = 3.46$, $MSE = 0.001$, $p<.05$. In both salient and non-salient target present trials, there was a simple main effect of set size (Fs(2,30)=12.8 and 5.8, respectively, MSE=0.001, ps<.01). This was not reliable in target absent trials, $F(2,30) = 2.4$, $MSE = 0.001$, p=.11. Importantly, in most cases the similarity scores were not significantly greater than the chance expectancy. Only in salient target trials with 12 objects ($t(15) = 3.18$) and target absent trials with 16 objects ($t(15) = 4.95$) was the intra-observer similarity reliable.

As I reported in Experiment 6, the similarity scores were much higher for the within-trial comparisons (Figure 8.7, bottom), indicating that there was consistency in the way different people viewed the search arrays. How did this consistency change across

experimental conditions? Effects of both set size, $F(2,30) = 29.01$, $MSE = 0.001$, $p<.001$, and trial type, $F(2,30) = 212.00$, $MSE = 0.001$, $p<.001$, were highly reliable. Similarity was greatest in the smallest set size ($t(15)=6.3$ and 7.2 for comparison with 12 and 16 items respectively, both $p<.001$), whilst the larger two did not differ. Within the different types of trial, all pairwise comparisons were reliable (all ts$(15)>2.8$, ps$<.05$), with scanpath similarity highest in salient target trials, followed by non-salient target trials. Target absent trials had the lowest inter-observer consistency. The interaction between trial type and set size was reliable, $F(4,60) = 6.28$, $MSE = 0.001$, $p<.001$. Simple main effects showed that in salient and non-salient target trials similarity varied with set size ($F(2,30)=11.8$ and 27.9, $MSE=0.001$, both ps$<.0005$), while in target absent trials it did not, $F(2,30)<1$. In all cases the scores were greater than chance (all $t(15) >14$).

8.4 General Discussion

These experiments looked at search for naturalistic objects—which vary across many feature dimensions—but in a controlled stimulus. Unlike the natural scenes used for the search tasks in Experiments 1, 4 and 5, these stimuli did not provide semantic or gist cues that would help find the target so top-down guidance can only have been provided by the knowledge of target appearance.

Consistent with the standard finding in visual search, there were effects of set size on the time to first fixate the target and the time to make a manual response, and these were very similar in both experiments. The presence of more items in the display meant that search was more difficult, presumably because, as in a feature conjunction task, items were serially inspected before the target could be located. In future experiments, it would be preferable to include a wider range of set sizes (it is not uncommon for visual search experiments to include, say, 4, 12 and 48 items, although these tend to be much smaller and simpler items). Despite this, the results were not inconsistent from the literature. Judging by the approximately linear increase in the time taken to fixate the target with additional distractors, it was possible to estimate a slope of around 30ms per additional item in Experiment 6. This is within the range of that reported by Wolfe (1998b) in his summary of many visual search experiments. It is important to note that, as this duration is much too short for additional saccades and

fixations (fixations averaged around 200ms) this may need to be explained by multiple, covert shifts, occurring within one fixation.

Despite the potential worry that people might not have correctly associated some of the items with their verbal descriptions, performance was very good in Experiment 6 (where only the description of the target was given), as well as in Experiment 7 (where a pictorial target was also presented). The additional instructions made a difference: manual RT and time to target were quicker in Experiment 7. Thus, the addition of visual information about the target allowed the task to be completed more efficiently, presumably by not having to inspect so many non-target objects. Wolfe et al. (2004) and Vickery et al. (2005) also reported that a more precise representation of a target in visual search made search more efficient, and here I extend this finding to more realistic objects. In the real world we often have incomplete knowledge about the appearance of an object that we are looking for (in fact given the multiplicity of views, lighting conditions and occlusions which we face any "template" would be unlikely to match exactly). The findings here are an example of a greater amount of top-down knowledge leading to greater filtering in attention. Consistent with this idea, the slope relating set size to the time to find the target was shallower in Experiment 7, showing that additional items in the display had less effect, which could be interpreted as an indication of greater parallel processing. An effect of more target information was not found in Experiment 1(a), where category search was predicted to be less efficient than instance search. In fact, there was no difference, even though instance search gave more specific details about target appearance. This may have been because in those stimuli the category exemplars were too similar to each other (and too different from the rest of the scene) to give any benefit.

There were some effects of target saliency on the search process, but these were minor. In the main, bottom-up saliency did not lead to faster search times or faster RTs. In both the experiments in this chapter saliency had an effect on the amplitude of the first saccade to reach the target. This is consistent with Experiment 1(b), and indicates that salient objects can attract attention and saccades from further away. This suggests that it is useful to look at saccade measures as well as just fixation time or location. On the

other hand, people were not reliably quicker at saccading to, or responding to, salient targets, even though they apparently came from further away. This discrepancy could be resolved if there was a latency cost in making a saccade to a salient target, although there seems no reason why this should have been the case. It may be best to remain cautious regarding saccadic amplitude, particularly as it was not affected by set size, as one would expect if this measure were sensitive to differences in search.

There was one further effect of saliency that is worth discussing. In Experiment 6, saliency interacted with set size: the increase in RT with more items to search was less steep for salient targets. A salient target was somewhat resistant to the addition of more distractors, and this is consistent with saliency affecting parallel, preattentive processing. Unfortunately, the same interaction was not found when looking at the time before fixating the target, and here saliency had no effect. So, although the RT data suggest that salient targets are more likely to "pop-out" amongst the larger set sizes than non-salient targets, in fact they were not found any quicker. The search tasks in Itti and Koch (2000) looked only at manual response time, but the experiments here show it is important to look at eye movement latencies as well. Experiment 7 also failed to find any effect of saliency on the speed of finding the target or responding that it was there. This is support for approaches that emphasise that bottom-up information is overridden or dominated by search.

The string-edit distance method for comparing scanpaths was well suited to the stimuli used here, as they were straightforwardly divisible into discrete objects. The analyses tested whether the sequence of objects which were fixated by participants was similar to (1) the order they were selected by the saliency map model; (2) the scanpath made when the same participant viewed other stimuli; and (3) the sequence by which other people viewed the same array. I argued in Chapters 4 and 5 that including several different comparisons allows an estimate of the different sources of scanpath similarity. In this task, the method looks at all the objects that are fixated during search, and not just the target, which gives a fuller picture of whether the saliency map model can predict the search path. In both experiments the results were the same: in the majority of cases the search scanpath was no more similar to the model's predictions than chance. When the

target was the most salient item, similarity was higher in most cases and this can be attributed to the fact that the task required participants to fixate at least one of the regions that the model had selected. It is more informative to look at those trials where there was not a target. In these trials, the model- and human-generated scanpaths were not highly similar, and only in one set size (out of six in the two experiments) was it reliably greater than that which would be expected from a purely random model of the search sequence.

So how was the sequence of objects chosen by the eye movement system? One possibility is that objects were selected at random. Another factor could have been a consistent scanning pattern imposed by the individual observer in all stimuli of this type, as might have been found if some people always started their search by fixating the top-left object and moving clockwise. However, the within-subject similarity was low and often not-reliably different from chance, particularly at small set sizes. In other words, knowing how a person moved their eyes previously in a stimulus with, say, 8 items, did not constrain how they would scan the next array with this number. In both experiments, similarity was higher with the largest set size, and it is probable that in these arrays, where there were more distractors and hence more noise, participants may have become more systematic with the order in which they fixated items. It remains to be seen how well this result would transfer to more realistic stimuli, although Experiment 2 suggested that people did show an idiosyncratic pattern when viewing scenes.

The third scanpath comparison looked at the similarity between different individuals viewing the same stimulus, and the results were promising. Scanning was not random and different individuals tended to look at similar objects in a similar order, even in trials where there was not a target. This could have been due to the fact that people started scanning in the same place (on any one trial). Alternatively it seems likely that people were fixating only those distractors which were most likely to be the target, for example because they shared similar features. This would narrow down the possible search paths and explain the reliable similarity found. Importantly it is this between-observers consistency that bottom-up models could potentially explain: information that is tied to the stimulus could determine the search path. However, as the comparisons with the saliency scanpath show, the saliency map model is unable to do this.

8.5 Conclusions and links forward

This chapter reported findings from two experiments involving searching for realistic objects amongst multiple distractors. Consistent with the visual search literature, adding more distractors meant that more objects had to be covertly attended or fixated before reaching the target, making search more difficult. Providing a picture of the target led to faster searches than providing just a verbal description. Target saliency did not have much of an effect on search, contrary to the predictions of a bottom-up model. Individuals did show consistency in the objects they fixated, as explored by comparing scanpaths, but this consistency was not accounted for by saliency and must instead be explained by top-down knowledge of target appearance.

In the salient targets used here, guidance by the task (i.e. target identity) coincided with guidance by saliency. What happens when salient items are unrelated distractors? The next chapter looks at the effects of salient distractors on search.

9 The effects of salient distractors in search

9.1 Introduction

The experiments discussed in previous chapters have shown some effects of model-predicted target saliency on performance in a search task. In Experiment 4, for example, salient target regions were found faster than non-salient target regions, although this was not found with target objects in Experiment 1. In most cases, participants were able to saccade to the target after only a few fixations. They were guided preattentively either by their knowledge of what the target looked like (Navalpakkam & Itti, 2005; Rao et al., 2002) or by cognitive expectations of where in a scene it might occur (Torralba et al., 2006). If these top-down guidance processes are able to completely override the attraction of salient regions then the selection of non-target regions (or distractors) should be determined only by the degree to which these regions resemble the target, and not by saliency. For example, if you are searching for a bottle of milk on a shelf you should be distracted by other bottles or white objects, and not by a very salient item which looks nothing like a bottle of milk. Chen and Zelinsky (2006) reported a result similar to this, showing that people almost never looked at a highly salient object (it was the only coloured item in a grayscale display) when it was not the target. If, however, saliency remains to be important in a natural visual search then one would expect the saliency of distractor objects to effect their fixation, and not just how much they coincide with the representation of the target.

This avenue was pursued in two experiments reported below. At first, it was hoped that a design featuring real objects in a natural scene would be possible. The results from this experiment (Experiment 8) were somewhat inconclusive, however. This can be partly attributed to the difficulty in controlling the layout of target and distractor in a real context. Experiment 9 used circular arrays of real objects and a brief viewing duration in an attempt to exert more control over the search process.

Broadly, the questions addressed are: whether a distractor affects searching for a target; whether this is moderated by its saliency; and how this depends on the location of this distractor relative to the target. Two further issues are worth introducing. The first is whether fixation of the distractor is necessary for an effect, or whether distractors might

prolong search even when they are not fixated. Relevant to this question is the "remote distractor effect" described by Walker, Deubel, Schneider and Findlay (1997). In these experiments the presence of another object in the display affected the latency of saccades to a target, even when the two were positioned far apart. The interpretation of this result is that the distractor increases competition within the saliency map that decides where the eyes will move (Findlay and Walker, 1999). In terms of locations, the potency of a remote distractor seems to depend on its proximity, not to the target, but to the current fixation (Walker et al., 1997). The stimuli in these experiments were dots on an empty screen. In the present experiments, one might expect a distractor, if it is fixated, to prolong search depending on whether it is near or far from the target. If it is not fixated, are effects found regardless of proximity to target?

A second point of interest is the landing position within an object. Previously, the experiments in this thesis have looked only at *whether* a target is fixated, and not at *where* within the target the eyes first land. If the saliency of objects around a target makes a difference, their effect might be seen in a shifting of covert attention or an alteration in the landing site. There is a large literature on saccade landing positions within words in reading that has identified the "preferred viewing location" (PVL) as being slightly to the left of the centre of the word (Rayner, 1979). The landing position has consequences for subsequent eye movements. For example, the "optimal viewing position" can describe the position in the word where identification is easiest and refixation is least likely, and this tends to be close to the PVL (O'Regan & Jacobs, 1992). Thus if the first fixation lands at the edge of a word an additional fixation on this word is more likely than if it lands in the centre, and conversely the fixation duration is likely to be shorter (O'Regan, Levy-Schoen, Pynte, & Brugaillere, 1984). Relatively few studies have looked at landing positions in objects. Henderson (1993) extended the findings from reading to a regular grid of four line drawings of objects that were inspected sequentially. Initial fixation position was distributed around the centre of the object, with more variability in the direction of the saccade than in the direction perpendicular to it. This position was also correlated with fixation duration and refixation probability; it was optimal to land near the centre. Experiment 9 allows landing positions within the target

object to be examined. This is less constrained than the sequential viewing in reading or Henderson's (1993) task so the results may be different. Is the landing position influenced by the saliency of the adjacent object?

9.2 Experiment 8: Salient distractors in a natural scene

9.2.1 Method

Participants

Twenty student volunteers with normal vision who had not taken part in the previous experiments took part.

Stimuli, apparatus and design

The eye-tracking apparatus was organised as in previous experiments using the EyeLink I, and stimuli were displayed in exactly the same way.

The stimuli were 96 colour photographs, showing household interiors involving furniture and other objects. Half of these contained a target that was a light blue glass marble ball approximately 1 cm in diameter. When photographed, the target was approximately 60 pixels square in size and was positioned to lie on an arc with a radius of 13˚ from the centre of the display. It was un-occluded and could lie anywhere on this arc (although due to the rectangular display some of the arc was outside the borders of the picture and so these positions were not used). Thirty-six of the pictures contained a salient distractor whose position was manipulated orthogonally, based on two factors: angular direction relative to the target (0, 90 and 180˚) and position relative to the centre of the screen (edge and centre). Figure 9.1b shows possible target and distractor locations. The resulting six conditions each contained six examples. Distractors were colourful and bright objects, approximately the same size as the target and chosen to be highly salient. Saliency maps were generated for each picture. All the distractors were one of the top three most salient parts of the scene, according to the model, while targets were never within the first 10 predicted fixations. Figure 9.1a shows a graphical representation of the saliency map for one stimulus.

Several steps were taken to prevent the distractor being consistently associated with the target. Firstly, the remaining twelve target-present stimuli featured a target with no distractor. Secondly, of the 48 target absent pictures, half contained a distractor. Finally, several different specific distractors appeared throughout both target present and absent trials. This meant that the only way to accurately find the target was to search for it, ignoring the presence and location of the distractor.

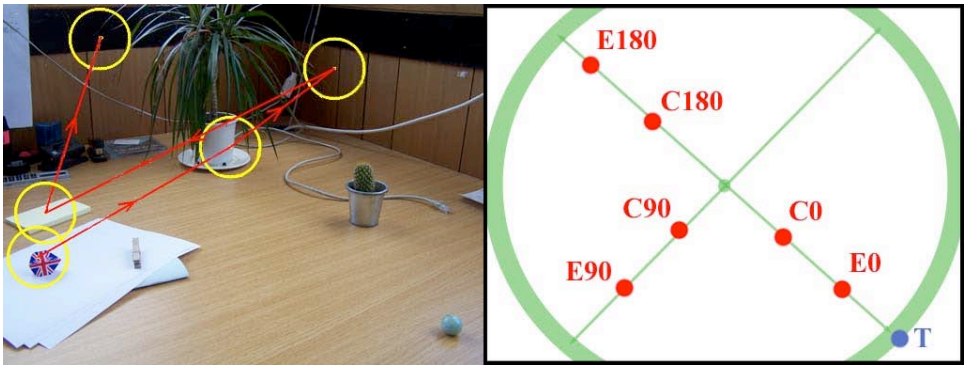


Figure 9.1. a) One of the stimuli used, with output from the saliency map model (left). Note that the target (bottom right) does not feature in the first five predicted fixations but the distinctive flag distractor is highly salient. b) A schematic showing possible object positions. The target (T) lay on or near the arc, and distractors varied in terms of position (edge, E, or centre, C) and angle relative to T (0˚, 90˚ or 180˚). Although only 90˚ locations are shown here, there was an equal chance of distractors in this condition being at 270˚ (or -90˚). The stimulus in a) is taken from the E90 condition.

A small set of similar practice pictures, containing target but no distractor, were prepared to show the participant what the target looked for in a real scene. In all the pictures, the target was kept non-salient by placing it in areas of the scene with similar colouring and intensity, making it blend in with its surrounding and therefore harder to spot. The two factors of angular direction and position of the distractor were manipulated within subjects to give 6 conditions (E0, E90, E180 and C0, C90, C180). Picture order was randomised.

Procedure

The participants were seated and the eye tracker calibrated as in previous experiments. They were told that their task was to search for a target and respond as to whether or not it was in the scene as quickly and accurately as possible. They were shown an example

picture of what the target looked like against a blank background. A short practice session showed several examples of target present and absent trials, followed by solutions where the target was highlighted with an arrow. This was then followed by instructions that emphasised that participants did not need to pay attention to other parts of the scene and should look only for the target. These were included to make sure that the participants, who had no reason to view these objects as important to the task, did not see any repetition of distractors as significant.

The experimental session consisted of all 96 pictures presented in a random order. Each picture was preceded by a drift correct marker and a fixation cross which ensured that fixation at picture onset was in the centre of the screen. Stimuli were offset by participants making a response via one of two keys on the keyboard to indicate whether or not the picture contained the target. There was no time constraint, and the next trial began with a drift correct marker immediately following the response.

9.2.2 Results

Two principal measures are of interest in this experiment. Firstly, reaction time data will identify whether task processes of planning eye-movements, scanning the scene, identifying the target and making a response were affected by the presence of a distractor. Secondly, eye movement measures will examine this in more detail to see whether saliency does indeed attract saccades and thus affect task performance. It should be noted that distractors might affect response measures without actually triggering eye-movements (for example by influencing covert attention, or as in the remote distractor effect). If there are no clear effects, is it the case that search can override the saliency map?

Both targets and distractors were defined by square regions of 2˚ and fixations within these regions were analysed. Trials where initial fixation was not within 1˚ of the centre were discarded, as were the minority that were answered incorrectly. Of principal interest here were target present trials, and a summary of the measures taken from these trials is given in Table 9.1. Participants tended to make around 5 or 6 fixations on average, which gives an indication of the difficulty of the task.

	Distractor position	Distractor angular direction (° relative to T) 0		90		180	
		M	*SEM*	*M*	*SEM*	*M*	*SEM*
Response time (all trials; ms)	**Edge**	1198	*56*	1027	*49*	1125	*64*
	Centre	1254	*77*	1288	*71*	1168	*65*
	Absent	1197	*33*				
Response time (where distractor is fixated; ms)	**Edge**	1128	*66*	1130	*69*	982	*59*
	Centre	1292	*90*	1445	*129*	1273	*66*
	Absent	1198	*56*				
Time to target (ms)	**Edge**	796	*40*	649	*30*	865	*83*
	Centre	934	*73*	862	*71*	770	*41*
	Absent	780	*28*				
Probability of distractor fixation	**Edge**	0.78	*0.04*	0.35	*0.04*	0.38	*0.05*
	Centre	0.69	*0.06*	0.46	*0.06*	0.44	*0.06*

Table 9.1. Measures taken from target present trials in Experiment 8, organised according to the position and angle of the distractor (where present).

Response time

This was the time from stimulus onset until the participant made their "yes" response, at which time the picture was removed and thus which also signified the total trial duration. A two-way (2 x 3), repeated-measures ANOVA with distractor position and angle as factors resulted in a main effect of position relative to the centre, $F(1,19) = 10.13$, $MSE = 42823$, $p = .005$, but no significant effect of angle, $F(2,38) = 2.27$, $MSE = 32770$, $p = .117$. There was no reliable interaction, $F(2,38) = 3.04$, $MSE = 48919$, $p = .060$. Overall, therefore, the only significant difference was that response times were shorter for pictures

containing edge distractors than for centre distractors. Figure 9.2 (top panel) shows the trial time for different distractor positions and angles.

Time to fixation

Another way of exploring the search process is to look at the time taken to move the eyes to the target for the first time. This should show a similar pattern to that for response time as in almost all cases first target fixation would be enough to classify the stimulus and respond accordingly. There was no significant main effect of either position, $F(1,19) = 3.90$, $MSE = 55769$, $p = 0.063$, or angle, $F(2,38) = 1.82$, $MSE = 66187$, $p = 0.176$. However, there was an interaction between the two factors, $F(2,38) = 4.44$, $MSE = 57946$, $p<0.05$, which was inspected for simple main effects. These showed that the locus of the effect of position was a difference between edge and centre trials at 90˚, $F(1,19) = 8.10$, $MSE = 55769$, $p<0.05$, whilst there was no effect at 0˚, $F(1,19) = 3.41$, $MSE = 55769$, $p=.081$, or at 180˚, $F(1,19)=1.62$, $MSE=55769$, $p=.22$.

Distractor fixation

The analyses above may have been complicated by the fact that the distractors were only fixated some of the time (around 45% on average). However, this varied between conditions. The proportion of trials where the distractor was fixated at least once showed no significant effect of distractor position relative to the centre, $F(1,19)<1$. The proportion of trials where the distractor was fixated did not vary between edge and centre positions. However, there was a significant effect of the angle relative to the target, $F(2,38)=44.13$, $MSE=0.033$, $p<.001$. Paired comparisons showed this effect was due to a much higher proportion of distractors being fixated when placed at 0˚ or in line with the target than when at either 90˚ ($t(19)=6.99$) or 180˚ ($t(19)=7.97$, both ps$<.001$). No other paired comparisons were significantly different and there was no significant interaction $F(2,38)=0.10$, $MSE=0.069$, $p=0.237$.

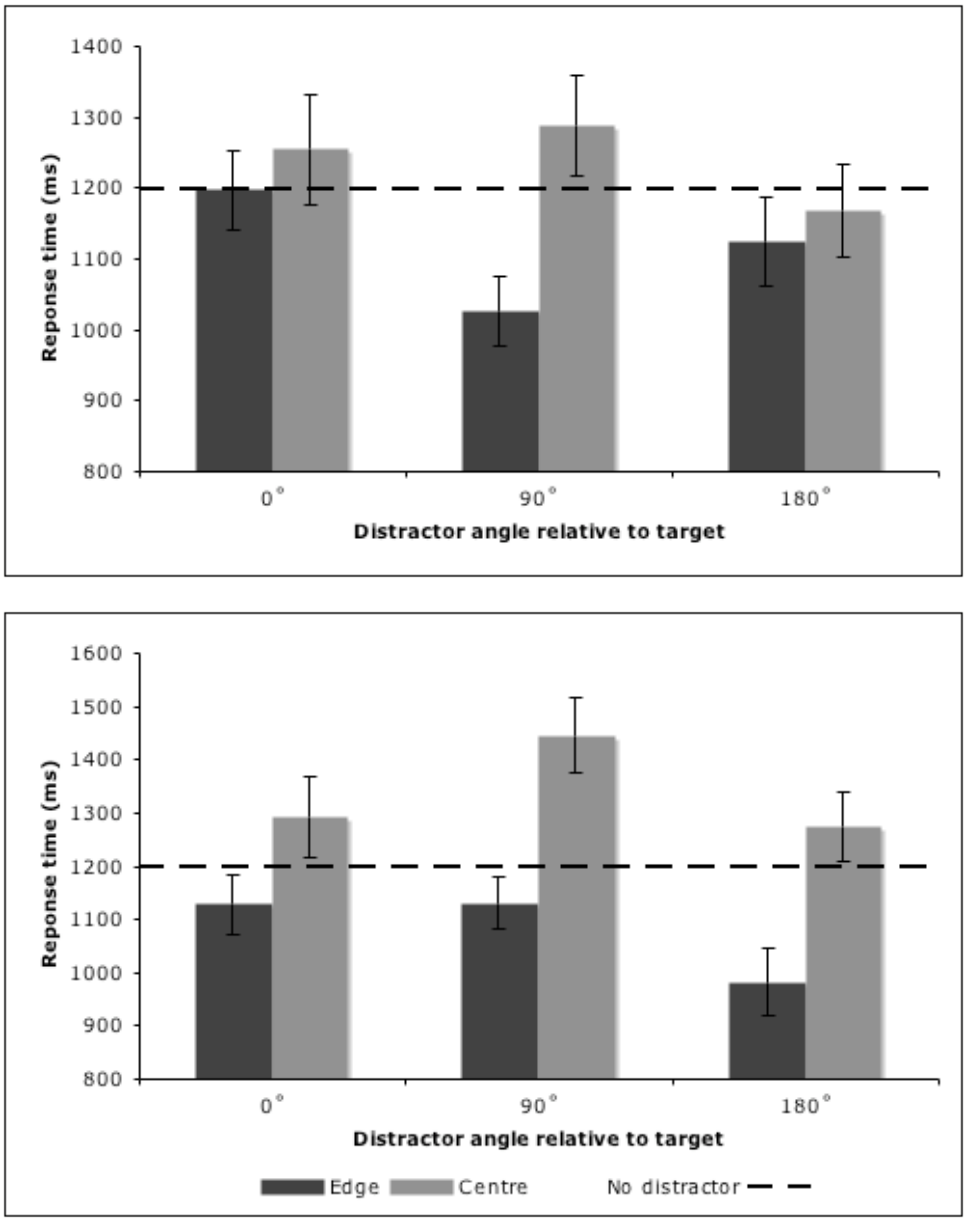


Figure 9.2. The mean trial time, equivalent to the response time, for all trials (top panel) and for only those trials where the distractor was fixated (bottom panel). Means are shown as a function of angle for edge and centre positions. Note that the mean trial time for trials without a distractor is included for comparison. Error bars give plus/minus one standard error of the mean.

Given that participants were able to disregard and leave distractors un-fixated, it might be that effects were different in trials where the distractor was fixated than in those where it was not. For this reason, the analysis of trial time was repeated but with the additional factor of distractor fixation (fixated v. un-fixated). This factor was significant, $F(1,19)=9.47$, $MSE=185758$, $p<.01$. Reaction times were significantly longer on trials

where the distractor was fixated (*M*=1208 ms) than when it was not (*M*=1037ms). The effect of position remained, $F(1,19)=6.01$, *MSE*=110333, $p<.05$, and angle was still non-reliable, $F(2,38)=1.20$, *MSE*=72164, $p=.16$ (see Figure 9.2, bottom panel). There were also several significant interactions in this 3-way ANOVA.

Firstly, whether the distractor was fixated interacted with the effects of both position, $F(1,19)=12.97$, *MSE*=106179, $p<.005$, and angle, $F(2,38)=5.69$, *MSE*=155124, $p<.01$. Simple main effects analysis showed that there was a significant difference between edge and centre distractors when fixated, $F(1,19)=17.9$, *MSE*=110333, $p<.001$ (such that central positions led to a higher mean reaction time than edge positions with 1337 ms and 1080 ms respectively). On non-fixated distractor trials there was no difference, $F(1.19)<1$. Angle had a large effect on reaction times both when the distractor was fixated, $F(2,38)=3.56$, *MSE*=72163, $p<.05$, and when it was not, $F(2,38)=10.59$, *MSE*=72163, $p<.001$.

Secondly, position interacted with angle, $F(2,38)=9.52$, *MSE*=70335, $p<.001$. Closer inspection showed that position only made a reliable difference at 180˚, $F(2,38)=16.6$, *MSE*=110333, $p<.001$, and not elsewhere, both $F(1,19)<1$. In addition to these interactions, there was a three-way interaction, $F(2,38)=5.58$, *MSE*=71596, $p=.008$, which might help characterise what difference is made to the effect of a distractor by fixating it. Overall, fixation was only reliable with central distractors at 0˚, $F(1,19)=12.68$, *MSE*=185757, $p<.005$, and those at 90˚, $F(1,19)=17.08$, *MSE*=185757, $p<0.001$.

9.2.3 Discussion

The results can be summarised as follows. Over all trials, responses were faster when targets were paired with edge distractors than when paired with centre distractors and sooner than those trials without a distractor. The time taken to fixate the target for the first time was also shorter for edge distractors, although this was not true when at 180˚. Distractors were actually fixated on only about half the trials, and distractor fixation was more likely when at 0˚ than at 90˚ or 180˚. Distractor fixation had an impact on response times, with fixation associated with slower responses. This also interacted with the effect of position in that when not fixated there was no reliable difference between trials with

edge distractors and those with centre distractors. The difference in trial times associated with different positions continued to be reliable in most cases, but was not so when distractors were fixated at 0˚. The angle of the distractor relative to the target had some effects on trial time but these were not clear-cut. Although distractors at 180˚ led to longer search times in some cases (over all trials, when distractors were central), when only fixated distractor trials were included they actually led to quicker responses.

This experiment provides some evidence that salient distractors disrupt the movement of the eyes in a search task, even though it would be more efficient to ignore them. Although this was the case on some occasions, perhaps what is most surprising is that on more than half of the trials the distractor was not fixated, even though it was highly salient according to the saliency map model. Presumably this is a case of search instructions overriding bottom-up attention, as has been seen in previous experiments in this thesis. There might be two ways in which a saliency map framework could incorporate search tasks. One is that the saliency map may be weighted to reflect target features. This could explain participants' failure to fixate distractors, as they may have been sufficiently dissimilar from the target to reduce their saliency to a point where they were no longer an important factor in determining fixations. An alternative way for the saliency map to be weighted would be based on probable target locations. Although a wide range of target placements was used, it might be the case that the edges of the picture were primed (as the target always occurred in these regions) or that other top-down assumptions about likely target locations (e.g. on surfaces) biased the search. However, this would suggest that edge distractors would be more likely to be fixated (regardless of their angular distance from the target) than those in the centre as at least some of these would become more salient due to being in these potentiated regions, but this was not the case. Instead, those distractors on the same line as the distractor in relation to the centre (those at 0˚) were much more likely to be fixated. This might be the case because the target, or at least some general information about the side or direction in which it lay, was gleaned from the periphery and only after this did salient distractors draw attention.

Search was slower when the distractor was fixated than when it was ignored, and there were several complex effects of its location. Can the explanations for these reveal something about the search process and its mechanism? When fixated, distractors located at the edge of the picture speeded up search. Although it is tempting to see this as a simple case of a distractor close to the target attracting attention and so facilitating search, this was not the case here as the effect of position was actually larger when the distractor was placed at 90˚ or 180˚ where it was further from the target than central positions and so should have increased search times the most. It might be that participants made an (incorrect) assumption that, after fixating an edge distractor the best place to look for a target was in a different quadrant and so search on these trials was quicker. Presumably this top-down guidance was not the case when the target was very close to the distractor (at edge 0˚) and so could be identified in the parafovea. However, although overall response time was shorter for edge trials, the time to fixate the target was longer for edge distractors at 180˚, suggesting that there must be more complex processing intervening between target fixation and response.

The angular direction of the target in relation to the distractor also had some effects on search efficiency. The fact that 0˚ distractors were more likely to be fixated has already been discussed. However, in terms of trial time and time to target, there were few clear patterns regarding angle. There was some evidence that distractors placed at 180˚ actually led to faster responses than those at other angles. This is a peculiar result although it might be explained when considering its interaction with position discussed above. An alternative explanation could involve something akin to inhibition of return when distractors were fixated at 0˚. If the target was not acquired from peripheral vision, participants might be less likely to look back here and this would prolong the average search time in comparison with those where distractors were on the other side of the screen. Again, though, the evidence from time to trial does not support this pattern.

To conclude, although salient objects do distract attention and make search longer this is not invariably the case and positional effects are not straightforward. A number of caveats should be noted regarding the design of this and future experiments. Firstly, although the object placement in this study was carefully controlled without

inspecting every fixation made it is hard to get an accurate picture of the courses taken by eye movements and their sequence. For example it may be that some of the fixations were "centre of gravity" fixations which landed between a target and a distractor and which were not analysed here. Secondly, as the saliency measured here by the model is ordinal, it should be noted that even in the distractor absent condition there would have been, by definition, a region that was most salient and so could be considered distracting. Without sacrificing the ecological validity of the stimuli it is impractical to control the saliency and location of every point in the scene and this may account for some of the variance in the above results. In order to provide a more controlled stimulus in which to investigate distractor saliency and fixation, Experiment 9 looked at much simpler arrays of coloured objects. Rather than manipulating both angle and eccentricity, this experiment focused on just the object next to the target (this was the distractor which was fixated most often in Experiment 8). This object was always the same distance away from the target, but it could intervene between the location where search started (similar to the central 0˚ condition in Experiment 8) or it could lie beyond it. Experiment 8 showed that a distractor that is relatively salient can be ignored but does have some effects. In the next experiment, the saliency of the adjacent object was varied, and the results investigate its effect on both general search efficiency and, more subtly, on the landing position within an object. Does a more salient distractor result in more disruption?

9.3 Experiment 9: Effects of an adjacent object in a speeded search task

9.3.1 Method

Participants

Sixteen student volunteers with normal or corrected-to-normal vision took part. All participants contributed to all the conditions. Following repeated calibration and drift correction errors and a high proportion of data loss, the eye movement data for one subject was excluded, although their response data is retained in the analysis.

Stimuli, design and apparatus

Given the difficulties in controlling target and distractor location in Experiment 8, this experiment used simpler arrays of objects. Figure 9.3 shows some examples of the stimuli used. The pictures were constructed by arranging eight objects on two imaginary concentric circles around the centre of the image. The whole display subtended 34˚ by 27˚. The inner and outer circles had radii of 6˚ and 11˚ respectively. The objects were taken from the same set of 30 as that used in Chapter 8. They were coloured items of food and drink and each one was resized and centred inside a region of constant size which was 3.5˚ square, although exact object size varied slightly (mean width=2.2˚, mean height=3.0˚). Two configurations were used in order to make the spatial locations less predictable (as depicted in Figure 9.3): one in which objects appeared on the horizontal and vertical axes, and one in which they were positioned on the oblique axes. The objects were presented on a constantly coloured background, which was one of 10 colours. As one of the features in the saliency map is colour contrast, altering the background colour could change the relative saliency of different objects. For example, a green pepper was less salient on a green background than on a red background.

The target was a single object that varied from trial to trial. The target probe consisted of this object presented at its normal size, centred on the screen (i.e. at the initial fixation position) and presented on a white background.

The object arrays were constructed using Adobe Photoshop 7.0 and this process was automated, allowing many unique stimuli to be created with random combinations of objects and backgrounds. No objects appeared more than once in the same array. The images were then screened using the saliency map model and the final stimuli chosen based on the following design.

Figure 9.3. Two stimuli from Experiment 9, demonstrating the two different configurations used. By manipulating the colour of the background, the saliency of objects can be changed. For example, the green pepper is salient with a grey background (top) but much less so with a green background (bottom).

Target eccentricity was defined as the circle on which the target was located (and therefore its distance from the centre) and could be inner or outer. The target was not salient according to the model; it was not selected within the first five shifts of attention made by the saliency model. Each object was featured as the target at least once. This study was particularly concerned with the effect of the object next to (and on the same vector from the centre) as the target. The object in this position was labelled the distractor. It should be noted that although the inner/outer label refers to *target* location, as the distractor was always in the adjacent position the converse label could be applied to

the distractor (i.e. if the target was inner, the distractor was outer, and vice versa). Distractor saliency was manipulated so that half the distractors were salient (the most salient object in the array according to the saliency model) and half were not salient (not featuring in the top five saliency rankings and therefore not different from the target). As the model measures relative saliency there is always a most salient object, but in the non-salient condition this object was not adjacent to the target. The selection of target and distractor was made with the following criteria: each target position was used equally and each object appeared equally often as target, distractor and elsewhere in the display. This ensured that no spatial locations or particular objects were any more important for the task than the others. The 2 x 2 design gave four conditions: inner (target) salient (distractor; IS), inner non-salient (IN), outer salient (OS) and outer non-salient (ON). The experimental stimuli set comprised 160 images, half of which contained the target. The 80 target present trials were divided equally into the four conditions.

Eye movements were recorded using the EyeLink II system and the standard set-up used elsewhere.

Procedure

The experiment began with a screen of written instructions followed by a short practice session of 10 trials, during which eye movements were not recorded, and which used images that were not presented again in the experiment. After the practice session the eye tracker was calibrated and the experiment proper began. Each trial started with the target object, presented for 2000ms. Following this, a drift correct marker appeared in the centre of the screen. Participants had to press a key on the keypad whilst fixating the marker and the experiment continued when the eye tracker confirmed this and corrected for any drift. The search array was then displayed and participants had to press one of two buttons to indicate the presence or absence of the target, as quickly as possible. If no response was made within 2000ms of array onset, a feedback screen indicated that the display had timed out and the trial was classed as incorrect. No other feedback was given. All participants saw the same 160 trials in a random order. There was the opportunity for a break after every 40 trials and the eye tracker was re-calibrated at this point.

9.3.2 Results

This study looked for effects of the relative saliency of an adjacent, distractor object on behaviour in a visual search task. The results were analysed in terms of the two independent variables: target eccentricity and distractor saliency. The results investigate the effects of these two variables on several measures.

First, task performance was assessed in terms of accuracy and manual reaction time; here the question is whether the saliency of the adjacent object determines if it interferes with the process of finding the target. If salient objects are more likely to draw attention and fixations than non-salient ones then they might distract more and therefore lengthen search or lead to more errors. This is further explored by looking at the eye movement data, specifically the time to first fixate the target, the initial saccade latency (how long before first moving the eyes) and the probability of fixating the distractor. A third set of measures looks deeper at the sequence of objects selected by the saliency model, and its correlation with the objects fixated when there is no target present. Is the probability of fixating an object, or of saccading directly to it, correlated with its saliency rank? Finally, the within-object landing position is measured to see whether people show a preferred viewing location and if this is affected by the saliency of the adjacent object.

Task performance

Trials were scored as errors when the response made was incorrect, or in rare cases when the trial timed out without any response being made. In all cases, the proportion of correct responses was close to ceiling (mean hits=93%, mean correct rejections=96%). Hits did not differ between the four conditions (IS=93%, IN=95%, OS=92%, ON=92%; No effect of eccentricity, $F(1,15)=1.05$, $MSE=0.0054$, $p=.32$; No effect of saliency and no interaction, both $F(1,15)<1$), although there was a slight trend, as one would expect, to be more accurate when the target was closer to the centre. The RT for correct trials showed an effect in the same direction (Figure 9.4). Eccentricity had a reliable effect, $F(1,15)=27.1$, $MSE=3620$, $p<.001$, as did saliency, $F(1,15)=8.6$, $MSE=2942$, $p=.01$. Participants were faster to respond to the target when it was closer to fixation (M= 761 and 839ms for inner and outer targets respectively), and they were slowed when they were next to a salient distractor (salient=820ms, non-salient=780ms). The interaction was

not significant, $F(1,15)=1.76$, $MSE=3426$, $p=.21$; distractor saliency had a similar effect on inner and outer targets. RT to target absent trials was slower (968ms compared to mean collapsed across target present trials, 800ms; $t(15)=9.5$, $p<.001$).

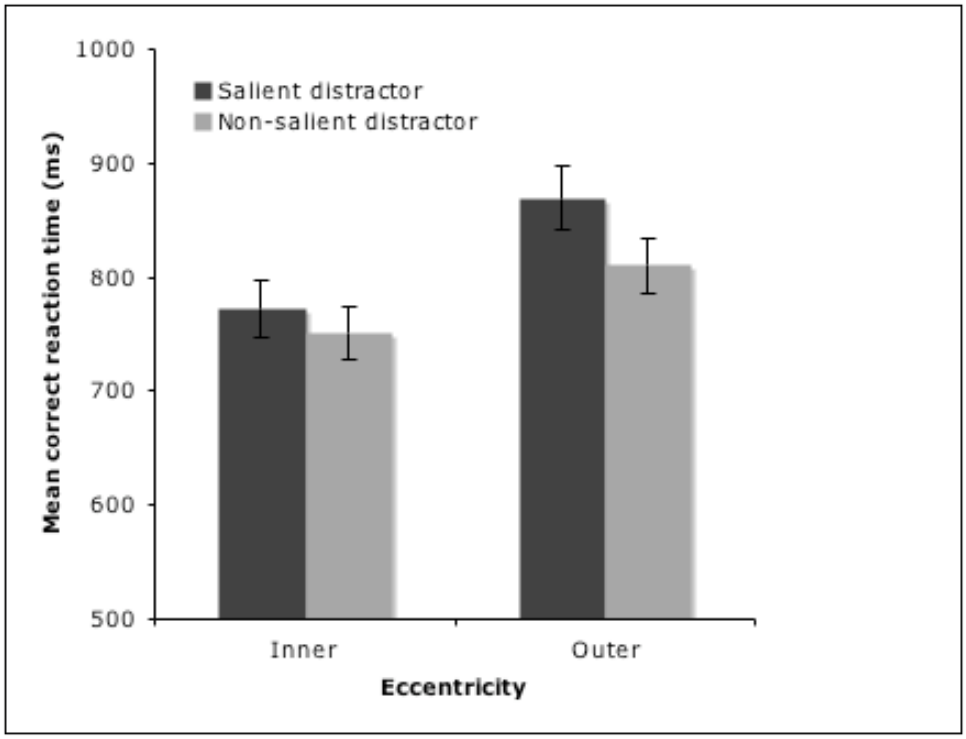


Figure 9.4. Mean reaction time, with standard error bars, as a function of eccentricity and distractor saliency.

Eye movements toward the target and distractor

The measure of time to fixate the target supplements RT by specifying search time, not including any time spent processing the target and deciding on the response. Trials that led to errors, and those where the target was not fixated at all were excluded from this measure. The pattern of results is similar to those in RT (Figure 9.5). There was a big effect of eccentricity, $F(1,14)=171.9$, $MSE=2089$, $p<.001$, with less time taken to fixate an inner target ($M=380$ms) than an outer target (534ms). Distractor saliency also had an effect, $F(1,14)=16.5$, $MSE=912$, $p=.001$, so that a salient distractor prolonged the time taken to fixate the target (473ms versus 441ms for non-salient distractors). These factors did not interact, $F(1,14)=2.3$, $MSE=1381$, $p=.15$.

Given that the saliency of an unrelated, adjacent object had an effect on searching for a target it is useful to ask whether this object was invariably fixated. If not,

was it fixated more often when salient? This might explain why it takes longer to reach the target: because the salient distractor captures fixations. The proportion of trials where the distractor was fixated was computed for each subject. The means showed that fixation of the distractor in the inner target conditions (i.e. when it was on the other side of the target) was very rare (mean percentage of trials, IS=5%, IN=1%). It occurred much more often when it intervened between the centre and the target (OS=38%, ON=42%). Repeated-measures ANOVA confirmed that there was a reliable effect of eccentricity, $F(1,14)=108.5$, $MSE=0.019$, $p<.001$, but no effect of distractor saliency, $F(1,14)<1$. Thus distractor fixation did not seem to fully determine the degree to which search was affected. There was a crossover interaction, which was marginally significant, $F(1,14)=4.1$, $MSE=0.0053$, $p=.06$. It is clear from the marginal means that this is because a salient distractor was more likely to be fixated than a non-salient one when it was on the outside of an inner target. When the target was on the outside the opposite was true: salient distractors were actually fixated less often.

A third eye movement measure, the initial saccade latency, was taken to see whether the distractors had an effect on the planning of the very first saccade. This measure was defined as the interval between stimulus onset and the start of the first saccade in a trial. It has been argued that saliency has the greatest effects early in a trial, so trials requiring saccades to a target whilst overriding a nearby salient distractor might take longer to program. There were no reliable effects, all $F(1,14)<1$. Initial saccade latency was fairly constant across conditions with a mean of 195ms.

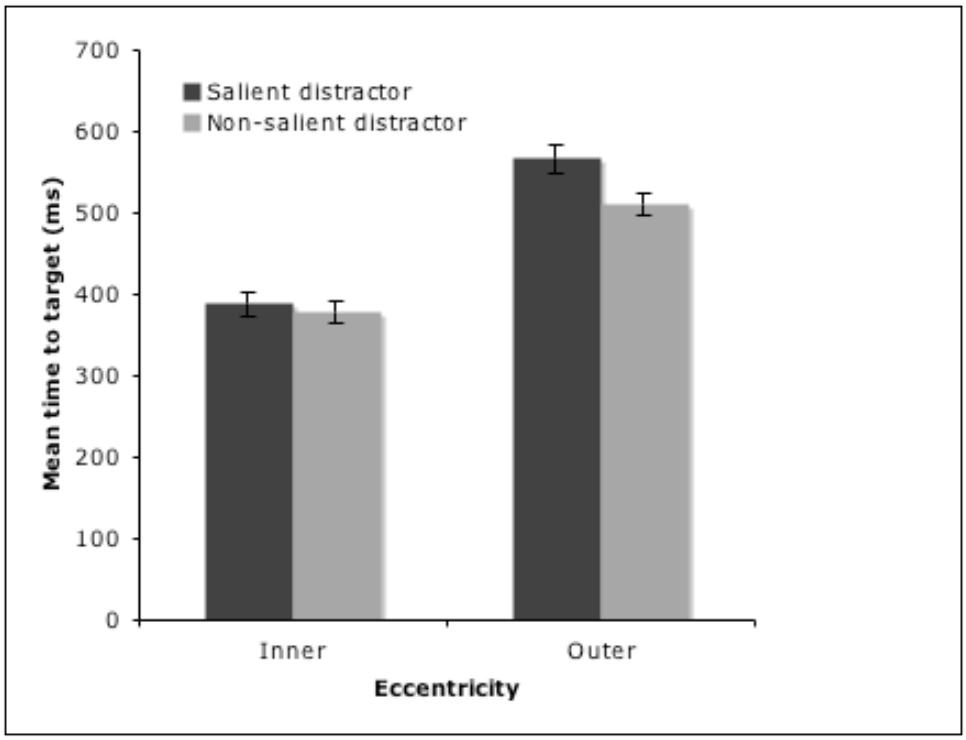


Figure 9.5. Mean time to fixate targets at different locations, and with different saliency distractors. Error bars show the standard error of the mean.

Saliency in target absent trials

The fixations made in target absent trials were inspected to look deeper at the influence of relative saliency of an object when no target signal was present. Given the 2000ms time limit there was a restriction on the number of fixations, and therefore the number of objects that could be inspected. An average of approximately 3.8 fixations (following the first) were made in target absent trials and at best half the objects could be fixated in the time limit. How were these objects selected? The saliency map model can predict the relative saliency of each object and therefore the order in which they should be fixated. To see how well this predicted the target absent search behaviour the probability of fixating an object at least once on a trial was plotted as a function of the object's saliency rank. This was equivalent to the proportion of occurrences of the object where it was fixated. The results are shown in Figure 9.6 and these were examined using repeated-measures ANOVA with factors of eccentricity and saliency rank.

Eccentricity had a large effect on the probability of inspection, $F(1,14)=149.2$, $MSE=0.018$, $p<.001$. As was the case with the distractor in target-present trials, objects in

target-absent trials were much more likely to be fixated when they were on the inner circle. Saliency rank also had an effect, $F(4,56)=6.5$, $MSE=0.004$, $p<.001$, but as can be seen from Figure 9.6 this was in the direction opposite to that expected. The most salient object in the array, ranked 1, was fixated somewhat less often than objects that were selected later by the model. A reliable interaction, $F(4,46)=4.1$, $MSE=0.0052$, p=.03, indicated that saliency rank only affected inspection probability in objects on the inner circle. This is probably due to a floor effect as objects in the outer positions were rarely fixated at all.

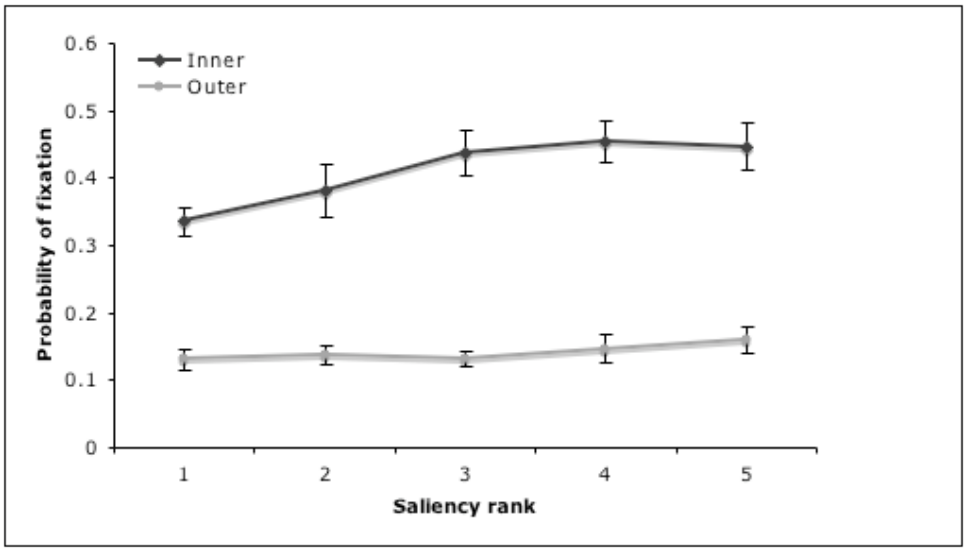


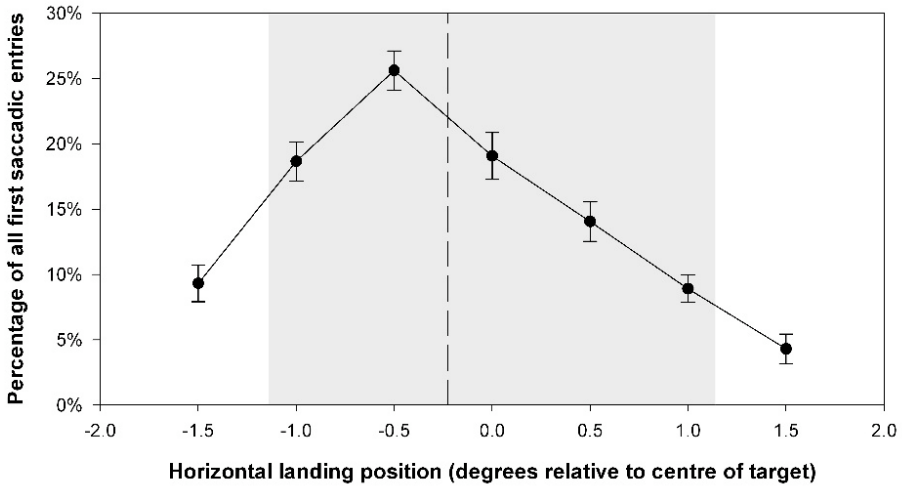
Figure 9.6. The probability of fixating an object in target-absent trials, according to its eccentricity and its model-predicted saliency rank. A probability of 1 would indicate an object with that rank and position was fixated every time it appeared in an array. Error bars show the standard error of the mean. Note that being selected later by the model did not lead to a lower probability of fixation.

Within-object landing positions

Landing positions were investigated by looking at the end point of the first saccade in a trial to enter the target region. In order to make the analysis more straightforward the saccade data was transformed in the following ways. The saccades were filtered according to amplitude: only saccades entering the target that were greater than 1˚ were included. This excluded small corrective saccades that came from just outside the target region. The landing position of the saccade was calculated relative to the centre of the target, so that a saccade that landed exactly on the centre of the target was considered to

have an offset of zero. To render all saccades and target positions comparable the offset from the target centre was separated into horizontal and vertical components, and the direction was coded relative to the centre of the search array. As this experiment was concerned with the effect of the distractor, offsets parallel to the vector between target and distractor were examined. In each case, offsets towards the centre (i.e. inside the object circle) were coded as negative and offsets towards the edge of the array (i.e. outside the object circle) were coded as positive.

Figure 9.7 shows the distribution of landing positions across all conditions, for horizontal and vertical offsets. They are plotted based on bins with a width of 0.5˚. Both chart a distribution that peaks on the inside of the target circle (between 0.25˚ and 0.75˚ away from the centre of the object). Between 20 and 25% of all saccades landed within these bounds, and slightly fewer landed in the central 0.5˚. The proportion of saccades decreases as the landing site moves further away from the target centre, with 59% and 52% of all saccades landing within 0.75˚ of the target centre, for horizontal and vertical offsets respectively. Beyond these regions the proportion of saccades decreases in both directions, with only around 10% of all landing sites over- or undershooting the target by more than 1.25˚, which would be beyond the edge of the target object. The mean (-0.23˚ and -0.19˚) and median offsets (-0.31˚ and -0.26˚ for horizontal and vertical respectively) confirm that people were most likely to land just before the target centre.



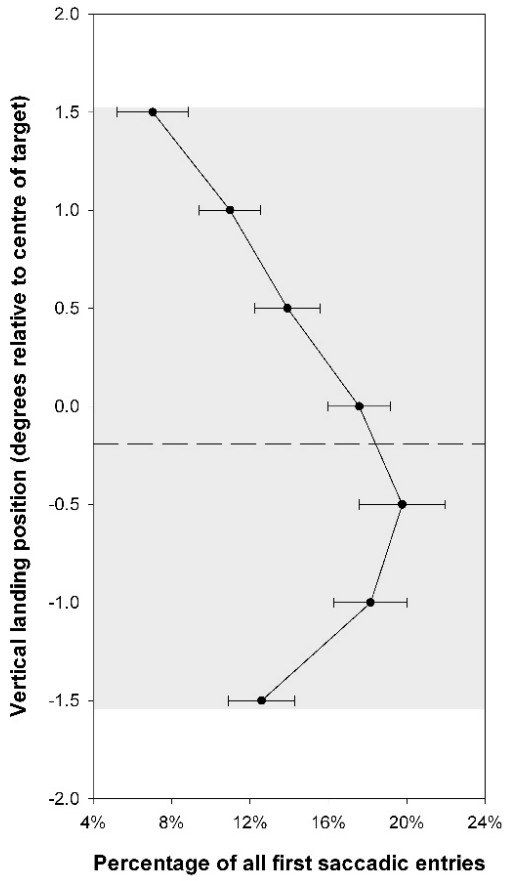


Figure 9.7. The distribution of landing positions within the target region for horizontal (top) and vertical (bottom) offsets. Data points show the mean proportion of all the first saccades that landed at this position (with standard error bars). The position axes show the mid-point of 0.5˚ bins. Negative offsets indicate undershoots which landed between the centre of the object and the centre of the whole array. The mean landing position is shown by a dashed line. The shaded area in each case represents the average dimensions of the target object within this region.

How do landing positions change with the saliency of the adjacent object? The arrangement of the stimuli allow for some clear predictions. If the presence of another object on the same line as the target makes a difference to the accuracy of the saccade then one might expect the average landing position of inner targets to shift outwards (towards the distractor). Conversely, outer targets, which have a distractor intervening between the centre and the target, might lead to a landing position shifted inwards. If saliency affects the extent to which an adjacent object attracts covert attention, this effect should be stronger for salient distractors. The distribution of first saccade landing positions is shown for each of the four conditions in Figures 9.8 (horizontal offsets) and 9.9 (vertical offsets). The distribution in each case is highly similar. To quantify any differences a repeated-measures ANOVA compared the mean landing position (dotted lines in Figures 9.8 and 9.9) as a function of eccentricity and saliency. Looking first at horizontal offsets, there were no reliable main effects (Eccentricity, $F(1,14)<1$; Saliency, $F(1,14)=1.1$, $MSE=0.057$, $p=.32$), and no interaction ($F(1,14)<1$). Vertical landing positions, however, showed a main effect of eccentricity, $F(1,14)=10.3$, $MSE=0.036$, $p<.01$, with outer targets leading to a more negative mean landing position (i.e. one landing closer to the centre of the array) than inner targets. Distractor saliency had no effect, and there was no interaction (both $F(1,14)<1$).

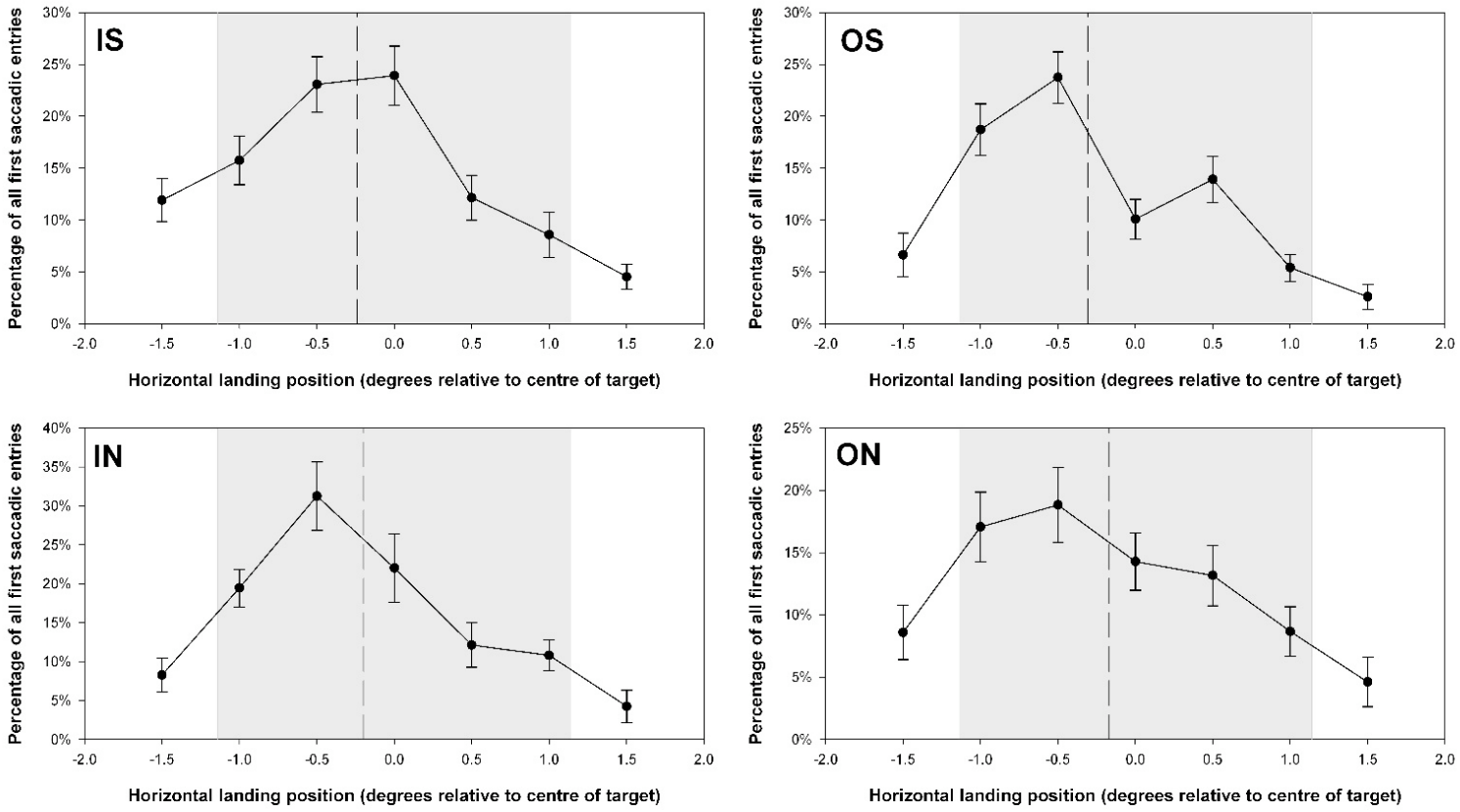


Figure 9.8. Horizontal landing position distributions for each of the four conditions.

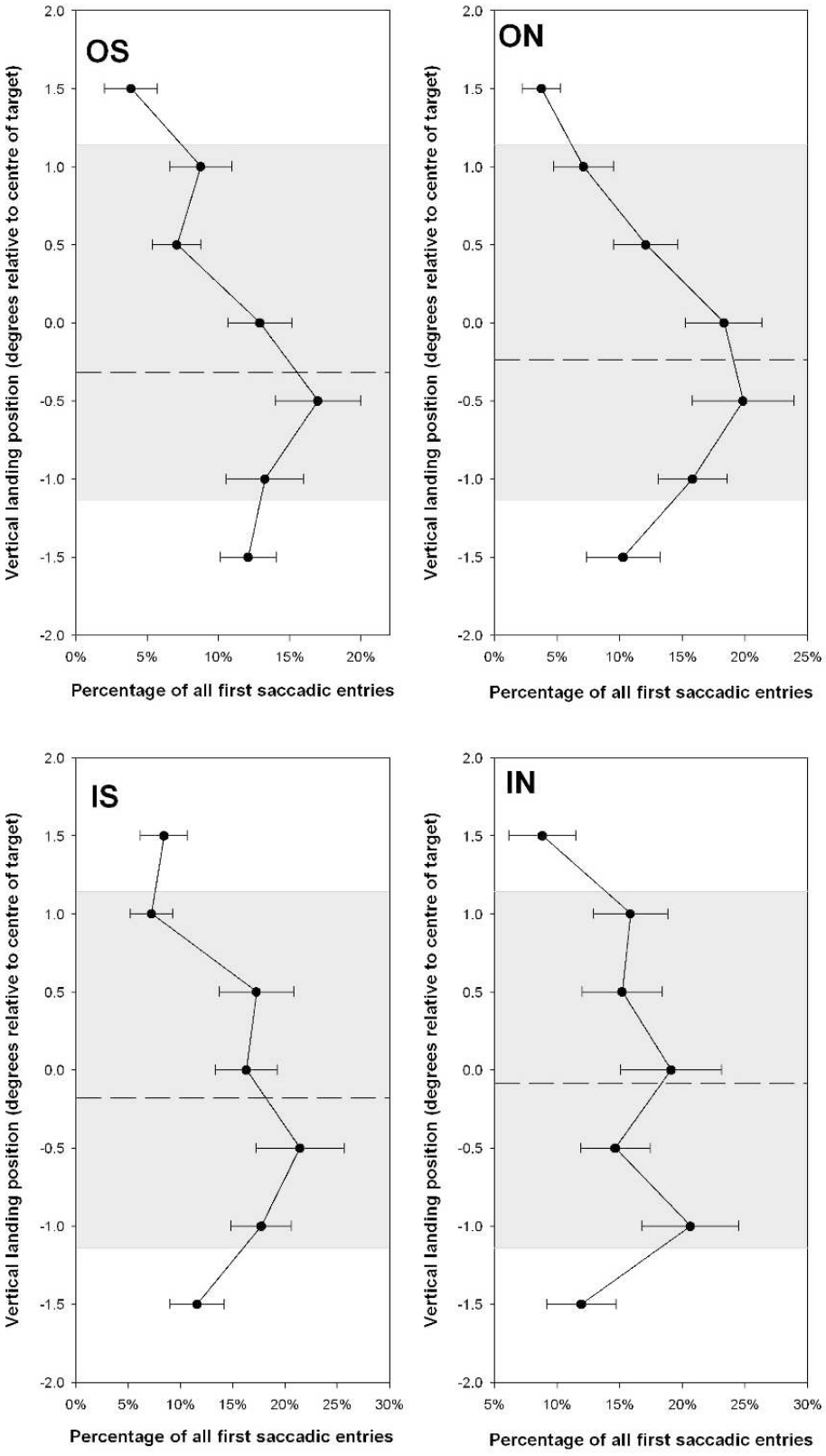


Figure 9.9. Vertical landing position distributions for each of the four conditions.

Effects of viewing position

The optimal viewing position (OVP) observed in reading single words is the fixation site that leads to the fastest processing. In objects, a central landing site is associated with the lowest refixation probability and the longest subsequent fixation duration (Henderson, 1993). To search for this effect of landing position on subsequent behaviour the Euclidian distance between target centre and first saccade landing position was compared with the likelihood of making at least one other fixation on the same object and the mean duration of the fixation following the saccade. If peripheral landing sites are non-optimal the target might need to be fixated again to be fully processed, and thus the current fixation might be terminated early to allow for a better position. Figure 9.10 gives the mean refixation probability and the mean fixation duration for each position. Consistent with predictions, these measures were not constant across landing positions. Refixations were relatively infrequent when the saccade landed within about 1.5˚ of the target centre, occurring about 30% of the time. When the saccade landed further away, a refixation became much more likely, so that a landing site more than 2˚ away (and therefore beyond the edge of the object) led to a refixation in 65% of cases. The mean fixation duration was less variable for all but the most distant position. Subsequent fixations at the very edge of the target region were much shorter than those that landed on the target. The effect of landing position on refixation probability was reliable, $F(3,42)=5.6$, $MSE=0.022$, $p<.005$, although this analysis did not include the most distant position as there were too few data points. The same ANOVA on the mean fixation duration at the first four positions was not reliable, $F(3,42)<1$.

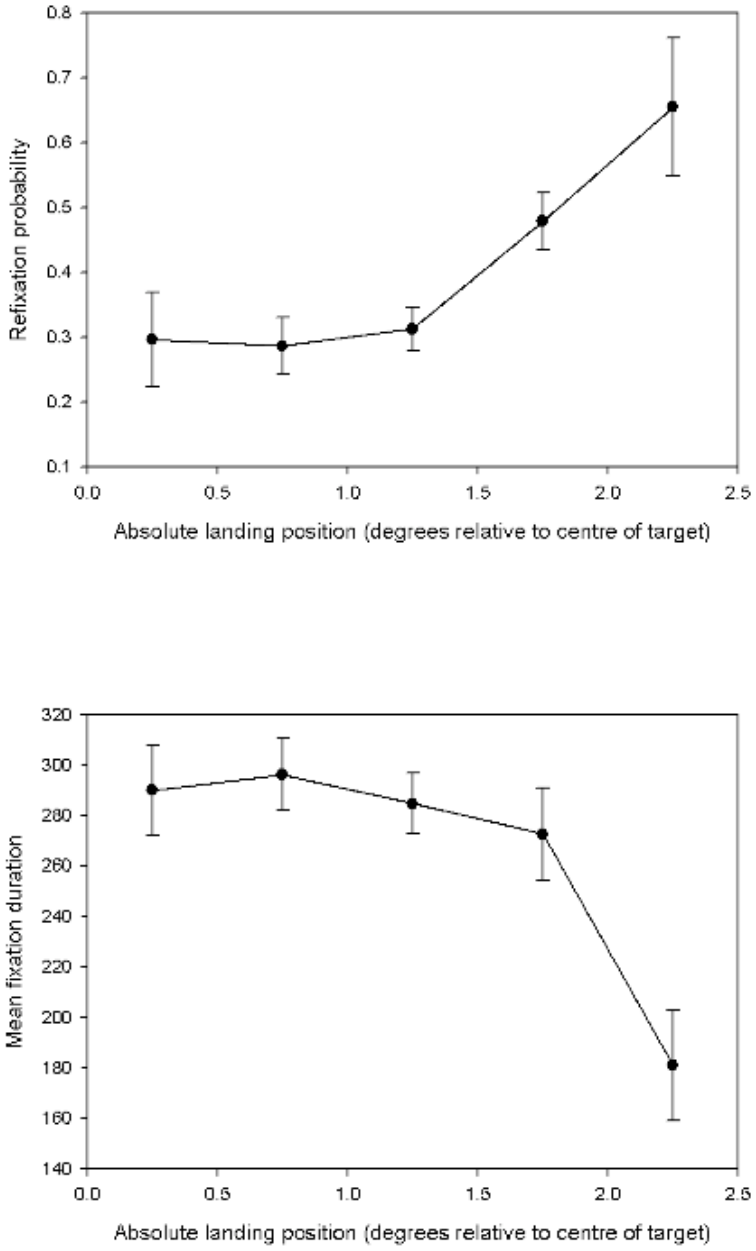


Figure 9.10. The optimal landing position within the target. The first entry saccades were binned according to their Euclidian distance from the target. Data points show the mean probability of refixation (top panel) and the mean duration of the subsequent fixation (bottom panel) for saccades in each bin. Error bars show the standard error of the mean.

9.3.3 Discussion

There were several interesting findings in this experiment. Object eccentricity had widespread effects on responses and eye movements: targets were found slower when they were on the outer circle (as shown by a reduced RT and time to target), and

distractors were less likely to be fixated when they were at this eccentricity. In target absent trials, eccentricity was an extremely reliable predictor of whether objects would be fixated. Of course, these effects are unsurprising given the visual system's reduced resolution in the periphery and the increased intervening area that needed to be searched (either overtly or covertly) before acquiring the target. The particularly low rate at which objects at the edge of the display in target absent trials were fixated indicates that it was relatively easy to reject objects as non-targets without fixating them. Given that each trial began in the centre, and that all objects were organised around this point, it is logical that participants would have avoided fixating the areas further away. Extensions to this experiment could perhaps control for eccentricity effects by enhancing the discriminability of objects further from the centre, or, if the intervening object in outer trials was responsible for the effect, using some arrays with no adjacent object.

Experiment 9 varied the saliency of the adjacent distractor in order to see whether a more salient distractor had more of an impact on search. There were several signs that this was the case: distractors that were highly salient according to the Itti and Koch (2000) model led to higher search times than non-salient objects. This supports the model and shows that bottom-up information can still have an effect on search, even when there were no systematic differences between the similarity of the target and distractor salient distractors were more likely to slow search. Cognitive override was not complete in this experiment.

There were some results that do not fit so neatly with the hypothesised effect of distractor saliency. Firstly, there was no effect on the initial saccade latency and hence no evidence that the presence of a salient distractor had an effect on the planning of the first saccade. This could be because saliency does not have effects very early in scene viewing, although this is counter to the observations of some researchers (Parkhurst et al., 2002; van Zoest et al., 2004). In Experiment 1(b) of this thesis the saliency of a target did have an effect on the first saccade. It might be that the relatively large distance between objects of interest and the initial fixation, or the number of fairly similar objects in the display can resolve this discrepancy. A bigger problem for a bottom-up approach in this task is the absence of a relationship between model-predicted object saliency and fixation

probability in target absent trials. Some might argue that saliency would be more likely to have an effect in these trials, as there was no accurate target signal available. Despite this, objects that were ranked highly salient by the model were no more likely to be fixated than those that were not. Whilst the search array was relatively less complex and varied than the natural scenes used in other experiments, Itti and Koch (2000) argue that their model is a useful predictor even in a simple feature/conjunction search for oriented lines. Assuming that the Itti and Koch model is a reliable estimate of bottom-up saliency then a top-down strategy must predict the fixation patterns. How were objects selected in these trials? Unlike natural search, the present results cannot be explained by a bias towards locations deemed more relevant for search: targets were equally likely to be positioned anywhere in the array so there was no contextual guidance in this way. Whilst it is possible that a consistent (or indeed a random) strategy was used, it seems likely that the objects selected were based on their similarity to the target. Several models have been proposed which could potentially simulate this with the complex objects here (Zelinsky et al., 2005; Rao et al., 2002).

The large number of trials, and the relatively constrained object locations, gave an opportunity to explore the within-object saccade landing positions. The results are fairly concordant with research in reading and objects that shows that people prefer to land close to the centre of the region of interest. The first saccades made to the target region here tended to land on the inside of the target, rather than exactly on the target centre. Making the reasonable assumption that participants were coming from the array centre (where search started) or from the other side of the display, this can be seen as a tendency to undershoot a saccade target, which is also an explanation posited for the left-of-centre PVL observed in reading (McConkie, Kerr, Reddix, & Zola, 1988). It may also have benefited participants to remain closer to the centre of the array as they were required to move back there for the start of the next trial.

If we identify the modal landing position (on the inside edge of the target) as a PVL we can ask whether this position was skewed by the presence of another object, particularly when it is salient. There was no reliable evidence for this, although for vertical offsets saccades to targets on the outer circle were skewed slightly towards the

distractor. However, this was not the case for horizontal offsets, and it might also be a "launch site effect"; saccades to outer targets would often have come from further away and so might be less accurate and skewed towards their starting point. The landing site in this study also affected subsequent eye movements, with saccades that did not land near the centre of the target being more likely to lead to a short fixation duration followed by a refixation. The implication of this is that the modal landing site is optimal for processing the target, and that when the saccade does not land here it is more efficient to terminate the current fixation and refixate than to continue trying to draw out the required information.

9.4 Conclusions and links forward

Whilst previous chapters have explored the saliency of search targets, this chapter looked at an object that was not useful for the task, and so presumably not part of the top-down set. This provides a different test of the saliency map model whereby two sorts of signals guiding search (one driven by knowledge of a target, and one driven by saliency independent of this knowledge) provide a conflict. The results were mixed, and on the one hand they show that saliency is not completely suppressed in visual search. Even when not fixated, a more salient distractor had more of an effect, hinting at some interesting effects on covert attention. On the other, the effects were rather small, particularly in Experiment 8, where distractors were rarely fixated and did not have an effect in all positions. This may be because in a natural scene there was more contextual guidance based on the constraints of where objects are likely to appear. In Experiment 9, which was better controlled, there were more effects of the distractor. However, when the target was not present the saliency map was demonstrably poor at predicting the order of fixation. In this case a search strategy based on target similarity probably dominated, although this should be tested in further experiments.

10 Investigating patterns in saccade direction

10.1 Introduction

Models of eye movements start from the basic finding that the places where people fixate are not random. In natural scenes, I have shown that eye movements look quickly towards the target (that is they are affected by search instructions). When no particular task constrains where to look, saliency models are better than chance at predicting fixation locations. In Chapter 5 I compared the predictions of the saliency map model to those of biased distributions to see if saliency adds anything, over and above image-independent, systematic biases in the way people look at scenes. One thing many bottom-up models have in common is that they assume, prior to any saliency computation, that all possible eye movements are equally likely. Given that visual acuity decreases rapidly with eccentricity, treating all retinotopic locations as equally likely to be fixated is problematic, and some researchers have addressed this (Parkhurst et al., 2002; Vincent et al., 2007). In other work it has been argued that systematic biases in which part of a display are fixated, in particular a central bias, should also be considered (Tatler et al., 2005). It is not always clear, however, whether there is a tendency to fixate centrally independent of the distribution of salient features, or whether fixations are biased towards the centre because salient objects are often located there. For example, the horizon in landscape photographs often provides a high contrast edge, which might attract attention in a bottom-up fashion.

In this chapter I investigate a related bias found in picture viewing: asymmetry in saccade direction. Several authors have reported that there are many more horizontal (leftwards or rightwards) than vertical saccades, and that there are even fewer oblique angle saccades. This is a general observation aside from the length of the saccade or its starting point, and the pattern has been seen in a variety of tasks and stimuli (H. F. Brandt, 1945; Crundall & Underwood, 1998; Gilchrist & Harvey, 2006). Why do people routinely move their eyes in this way when viewing scenes? There are several possibilities.

Firstly, in what I will call the *oculomotor* explanation, the distribution of saccade directions might be due to dominance of the muscle or neural apparatus that triggers

horizontal shifts of the eyes, regardless of the stimulus being viewed. Although most physiological research has concentrated on horizontal saccades there is some evidence that vertical and oblique saccades are slower and exhibit more curvature than horizontal saccades (Becker, 1991; Becker & Jurgens, 1990; Collewijn, Erkelens, & Steinman, 1988). This would support an image-independent bias for horizontal saccades, and such a bias would be expected across different stimuli.

Secondly, and of particular relevance to this thesis, an *image-characteristics* explanation could explain this bias in terms of the distribution of salient features in pictures and natural scenes. The horizon often features in outdoor scenes and this is normally marked by a high-contrast edge between dark ground and lighter sky. It might also be the case that the semantically important objects in the scene (people, cars, buildings) are found near this horizon. Photographs of the natural environment are usually composed with these objects in the centre (in fact a beginners' photography heuristic suggests that the horizon should be around two-thirds of the way up the picture). This non-uniform distribution of salient features has been identified as a confound in studies which show a correlation between saliency and fixations (Tatler et al., 2005). If people are reflexively drawn to regions of high contrast or high saliency in the periphery (as suggested by saliency map models), and if these regions tend to be positioned horizontally from each other, then this would cause a predominance of horizontal saccades. It has also been noted that natural and manmade scenes tend to have more horizontally and vertically oriented contours than oblique ones (Coppola, Purves, McCoy, & Purves, 1998), which could be an image-based determinant of saccade direction.

Alternatively, a *learned* account could predict a horizontal bias based on our experience with pictures and the environment. By this account, horizontal saccades are not favoured automatically by neurophysiology or caused by relatively automatic orienting to salient objects on the midline but rather learned over time and initiated top-down. Following multiple experiences with scenes where important information (both visually salient and semantically interesting) is located on the horizon humans learn to move their eyes in this way, in order to maximize the details observed in the fewest saccades. This learning might be subject to cultural and experiential differences, for

example, in terms of reading habits. Consistent with this Abed (Abed, 1991) reported differences in scanning direction between Western, Middle-Eastern and East Asian participants looking at simple dot patterns. While Western readers made more shifts moving from left to right, Arabic readers were more likely to show the opposite pattern. East Asian participants showed a 1:1 ratio of horizontal to vertical saccades unlike the 2:1 ratio seen in other readers. In a driving experiment, Western drivers showed a ratio which was closer to 4:1, though interestingly there was no difference between experts and novices despite the former presumably having more experience with the layout of the road (Crundall & Underwood, 1998; Crundall, Underwood, & Chapman, 2002).

A more specific instance of this would be a *learned layout* explanation. In this more flexible account, basic cues about the layout of the scene might influence the likelihood of moving along each axis. As discussed in Chapter 1 the overall gist of a picture (for example whether it is outdoors or indoors) can be acquired very rapidly (Biederman, Rabinowitz, Glass, & Stacy, 1974; Potter, 1976;). Coarse information gathered from the first glimpse might also include simple knowledge about the location of the horizon or the overall structure. In a series of experiments, Sanocki (2003) showed that briefly shown scenes can prime spatial layout, and that this priming affects subsequent perception. This supports the proposition that scenic layout, and not just scene category, is represented following an initial glimpse. This knowledge could affect which way the eyes are likely to move.

A final possibility is that the biases in saccade direction are *display-specific*, an artefact of laboratory-based eye tracking studies that present scenes on a computer monitor. These monitors are normally wider than they are high, and pictures are often presented in the landscape orientation filling the screen. Thus it may be more efficient to move horizontally than vertically in order to cover the whole area. Experiments often cue attention with a fixation cross in the centre of the screen at the start of a trial. In this case there is more information to the left and right of fixation than above or below, and this continues to be the case with an asymmetric display. In addition, older studies often suffered from large tracking errors, which tended to be greater in vertical saccades than in horizontal ones (Yee et al., 1985). In the real world, biases in saccade direction might be

different, although Crundall and Underwood (1998) found the bias occurred with real roads. Of course the human field of view with two eyes is also asymmetric, spanning up to 180˚ horizontally and only around 90˚ vertically, and this is extended further with head movements that move further horizontally than vertically. These head movements affect the way the eyes move in the real world (Pelz, Hayhoe, & Loeber, 2001; Smeets, Hayhoe, & Ballard, 1996).

The experiments reported here investigate the distribution of saccade directions in an attempt to distinguish between these explanations. The accounts summarised above are not mutually exclusive and several of them might combine in natural vision. For example when combined with a central starting position in the laboratory bottom-up information might drive saccade direction. The environment in which humans find themselves is likely to have shaped our physiology in terms of field-of-view and head and eye musculature. Studies of non-human animals demonstrate that the environment, and in particular the prominence of the horizon in the natural habitat, affects the organisation of the retina (Hughes, 1977). How flexible is the human tendency to move the eyes in a certain way, and when in viewing does this tendency emerge?

To answer these questions in Experiments 10(a) and 10(b) I rotated natural images from the horizontal whilst recording the eye movements made in a simple scene understanding task. Various elements of the display were controlled in order to remove artefacts of the laboratory set up. If the pattern of saccade directions is due to oculomotor factors or long-term learning then it should be insensitive to trial-by-trial variations in scene orientation. On the other hand, if the pattern changes with scene rotation the bias must arise from changes in the distribution of salient features or early recognition of the scene layout. A special case concerns scenes that are rotated 180˚ and are therefore upside-down. Inverted scenes will preserve any clustering of features around the horizon but scene inversion might disrupt gist acquisition and scene recognition. In this case it will be informative to discover whether the saccades resemble those in normal, correctly oriented scenes.

10.2 Experiment 10(a): landscape orientation in a sentence verification task

10.2.1 Method

Participants

Thirteen student volunteers from the University of Nottingham with normal or corrected-to-normal vision took part for payment. All were naïve to the purpose of the experiment and gave their full, informed consent to participate.

Stimuli, design and apparatus

This experiment used landscape photographs that all had a visible horizon and thus tended to have a large contrast boundary running horizontally through the middle third of the picture. Forty colour photographs of landscapes and outdoor scenes were chosen from a commercially available CD-ROM or taken with a high-resolution digital camera. Some of these had previously been used in Experiments 2-4. To create the normally oriented set these pictures were then cropped into square images of 768 by 768 pixels around their centre (see Figure 10.1). Making all the images square helped to control for any effect a rectangular display might have on saccade direction, and gave a consistent frame of reference across all trials. The original full-size images were then rotated by 45, 90, 135 and 180˚ using photographic manipulation software. Cropping these into squares gave the rotated stimuli sets. Preparing the stimuli in this way ensured that the borders of the picture were square in each case, whilst the horizon was rotated. The final stimuli were composed of five rotation sets (0, 45, 90,135 and 180˚) with eight pictures in each set. As this experiment was mainly concerned with the axis on which the picture was aligned, the pictures in the 45, 90 and 135˚ sets contained pictures that were rotated both clockwise and anti-clockwise from the horizontal (so that pictures in the 45˚ set were equally likely to contain an orientation aligned at +45˚ and +225˚).

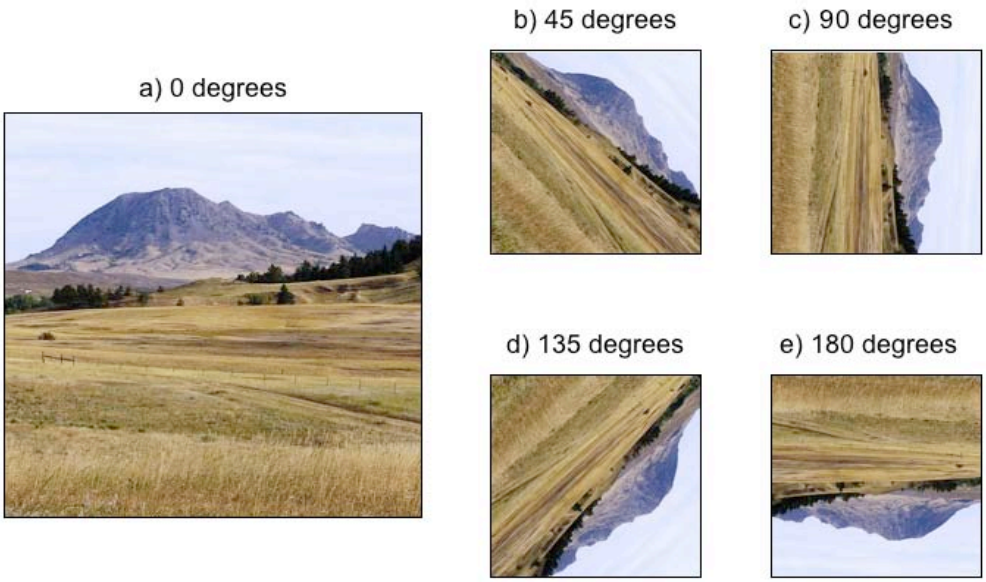


Figure 10.1. An example stimulus from the experiment. The final rotation sets also contained images at four different rotations (b-e).

The pictures were displayed on a monitor and the eyes were tracked using the standard EyeLink II set-up. It is important to note that spatial resolution was high for both horizontal and vertical movements (error was less than 0.5˚ in each direction), and where this was not the case the calibration was repeated until this level of accuracy was achieved. A chin rest was used and participants were instructed not to move their head during trials. A head camera on the eye tracker also monitored head movements but these occurred very infrequently. When these occurred participants were reminded to keep their head still, the trial number was recorded, and data from this trial was excluded from the results. Responses were entered using a gamepad.

Procedure

Following calibration, a short practice example was shown prior to the experimental trials. Each trial (see Figure 10.2) began with a drift correct dot, which re-aligned the calibration and also had the effect of forcing the participant to start scanning at a particular place in the image. In order to avoid artefactual effects on saccade direction of a central starting point, the drift correct location was varied from trial to trial and

appeared in one of the four corners of the display (approximately 4˚ from the edge of the monitor) and just outside the boundary of the square stimulus. Each starting point was used equally within each rotation set but was otherwise random. After fixation of this point was confirmed the picture was displayed for 2.5 seconds.

Participants were instructed to take in as much information as possible, and each picture was followed by a written sentence presented on the screen. Participants were required to verify the truth of the sentence in terms of the picture they had just inspected, and they did this by pressing one of two keys on a keypad to indicate true or false. All sentences were active declaratives referring to the identity or location of objects or scene features (see Figure 10.2 for example). Each rotation set was associated with an equal number of true and false sentences. The trial ended with the participant's response, and then the next trial began. All forty stimuli were presented in a randomised order that was unique for each subject. Eye movements were recorded while the picture was presented, and image onset and offset times were also written to the data file to ensure that no eye movements from reading the sentence (which would largely be horizontal) were included.

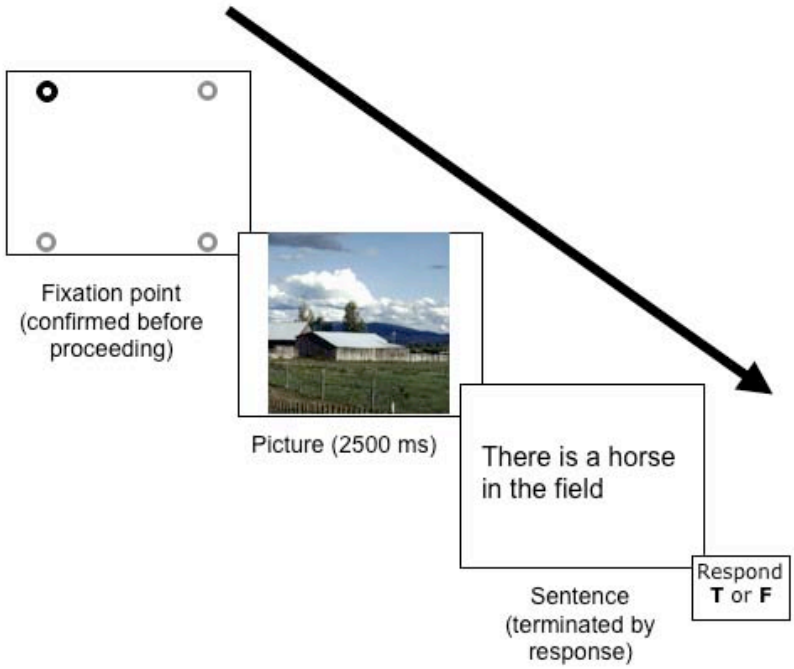


Figure 10.2. The procedure for one trial. Fixation started at one of four points (grey circles). The picture was then shown for a fixed period before a sentence verification checked picture understanding.

10.2.2 Results

Participants were quite accurate at the sentence verification, showing that they were able to do the task (mean proportion correct 0.82). The remainder of the results aim to characterise the saccades made whilst encoding the pictures for the sentence task.

Saccade direction

The angular direction of all saccades made whilst inspecting the pictures was recorded. The first saccade in each trial started at the experimenter-controlled drift-correct location and as such it was examined separately. Inspection of these saccades and their landing points showed that in the vast majority of cases the direction and amplitude was determined by the start point, with a general tendency for the saccade to move to the centre of the screen (see Figure 10.3). As a result all further analyses looked only at subsequent saccades (N=4320).

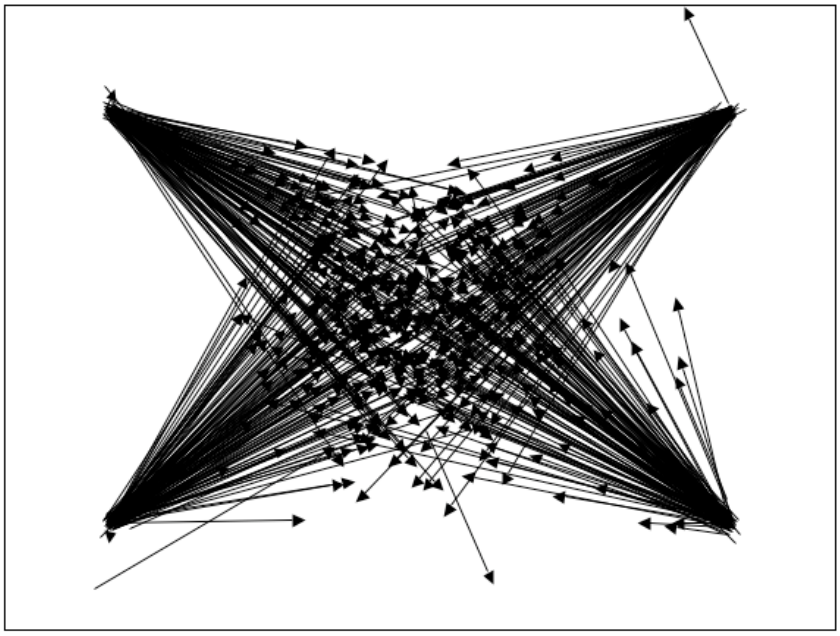


Figure 10.3. The first saccade in each trial, across all participants. Trials started in one of four locations, but the first saccade was almost always made towards the centre.

To describe the overall pattern of saccade directions associated with each picture orientation, the following procedure was followed. First the small number of saccades shorter than 1˚ were removed, in order to exclude readjustive and microsaccades. This removed less than 5% of the data. All possible directions were then divided into 36 bins

of 10˚ each. To remain consistent with the labels for picture orientation, these bins were numbered clockwise from the horizontal, starting at 0 for saccades that lay between 350 and 0˚ (i.e. almost exactly leftwards). The proportion of saccades in each bin was then plotted in a polar plot. These plots are shown in Figure 10.4 separately for each picture orientation (a-e). Looking first at the normally oriented condition (a), it is clear that there is a strong horizontal bias, with more than twice as many saccades in the 0 and 180 degree bins than in the 90 and 270˚ bins. There are also more saccades in the vertical than in the oblique. The peak of saccades on the horizontal axis is particularly pronounced for rightwards (180˚) saccades. Figures 10.4(b-d) show clearly that this pattern changes with the orientation of the picture, so that more saccades are made in the axis in which the horizon of the picture lies; saccade direction is effected by picture orientation. Figure 10.4(f) combines all orientation conditions by rotating the direction bins so that the charts horizontal (0˚) is aligned with the original orientation of the picture, showing a relative horizontal bias across conditions.

In order to make some clearer comparisons, the same data were divided into four bins according to the closest axis (horizontal, 45/ 225˚, vertical and 135/315˚). The bins were defined using all eight directions (four cardinal and four oblique) +/- 22.5˚, so that, for example, the 45 / 225˚ bin contained all saccades greater than or equal to 22.5˚ and less than 67.5˚, along with all those greater than or equal to 202.5˚ and less than 247.5˚. Whilst this loses any asymmetries in terms of left or rightwards saccades, it allows for a more straightforward analysis. I confirmed that most of the variation in the distributions is symmetrical by comparing the frequency of leftward (<90˚ or >270˚) and rightward (>90˚ and <270˚) saccades. This is equivalent to comparing the left and right sides of Figure 10.4(f), and there was no difference ($t(12)<1$) and therefore no evidence of any asymmetry. The frequency of saccades within each axis was computed, for each subject, in each orientation condition. Figure 10.5 shows these data as a proportion of the total number of saccades made in each condition. The data were compared using two-way (4 directions by 5 picture orientations) repeated-measures ANOVA. The oblique picture orientations contained pictures rotated both clockwise and anti-clockwise. Planned *t* test

comparisons then compared the frequency of saccades in the same axis as the picture orientation with the mean of the other three axes.

Across picture conditions there was a reliable effect of direction on saccade frequency, $F(3,36)=3.97$, $MSE=36.2$, $p<.05$, with fewer saccades being made in the 135˚ axis than in the horizontal axis (post hoc t test with Bonferonni correction, $t(12)=3.58$, $p<.05$). No other comparisons were reliable. There was no main effect of picture orientation, $F(4,48)=2.11$, $MSE=1.77$, $p=.09$, indicating that roughly the same number of saccades was made regardless of how the picture was rotated. Most importantly, the change in the complete direction distribution is shown in this analysis by an interaction between orientation and axis, $F(12,144)=21.25$, $MSE=20.64$, $p<.001$. It can be seen from Figure 10.5 that this interaction is due to the greatest proportion of saccades in any one condition being made close to the horizon's axis. This was confirmed with planned comparisons which compared the mean frequency of saccades in the axis where the picture's horizon was located with the mean frequency of saccades elsewhere, collapsed across the remaining levels. In all cases the comparison was highly reliable (at 0˚ orientation, $t(12)=6.61$, $p<.001$; at 45˚ orientation, $t(12)=5.45$, $p<.001$; at 90˚ orientation, $t(12)=5.72$, $p<.001$; at 135˚ orientation, $t(12)=3.31$, $p<.01$; at 180˚ orientation, $t(12)=5.41$, $p<.001$). Thus whichever way the picture was oriented there were more saccades in the axis corresponding to the picture's (horizontal) orientation than elsewhere. Of particular interest is the comparison between pictures oriented normally (0˚) and those inverted (180˚) and Figure 10.5 shows that the distribution of saccades in the four axes is highly similar between the two.

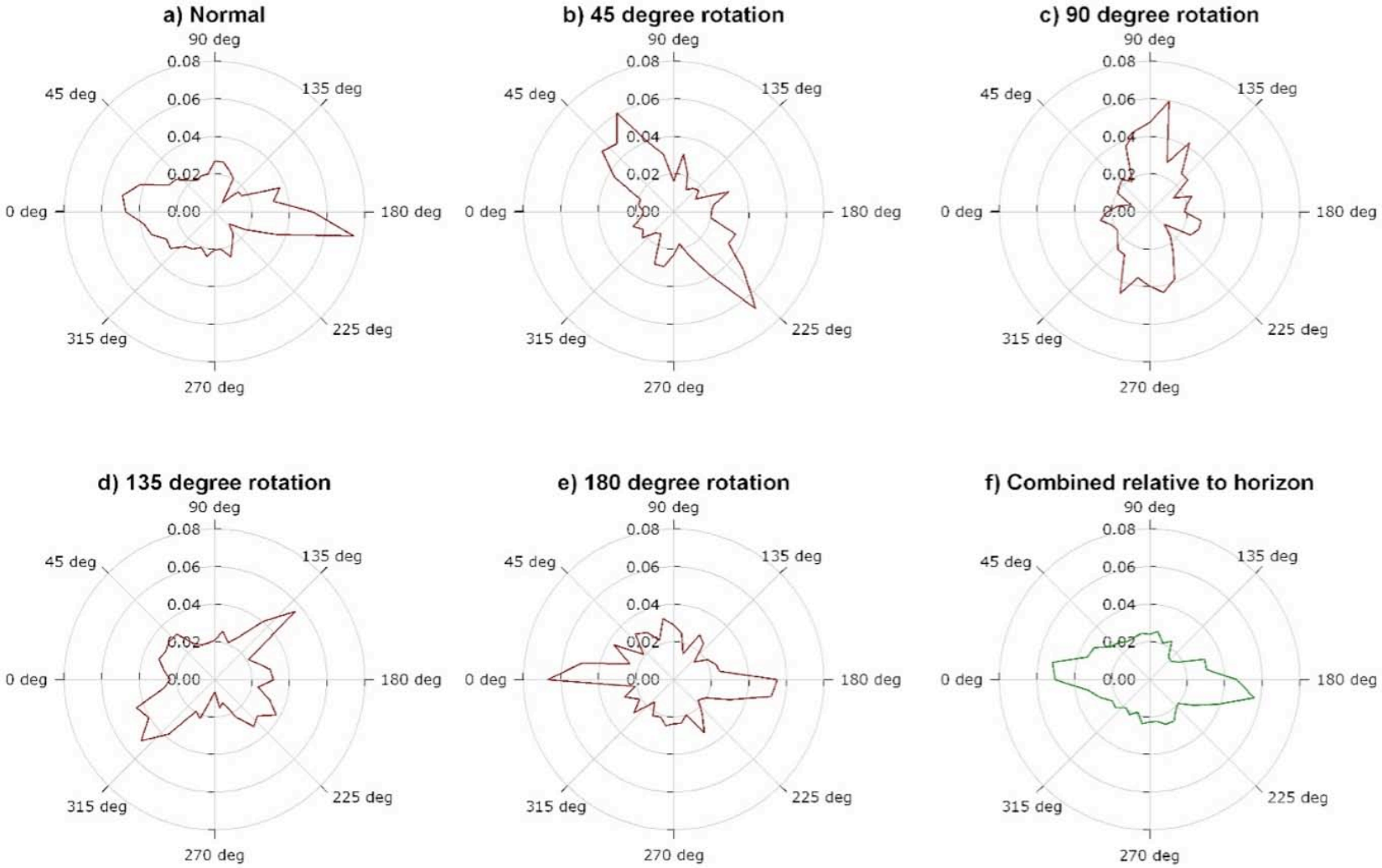


Figure 10.4. Radial histograms for all saccades (excluding the first). Each plot shows the proportion of saccades (y axis) in each of 36 direction bins. The majority of saccades occur in the axis of the horizon.

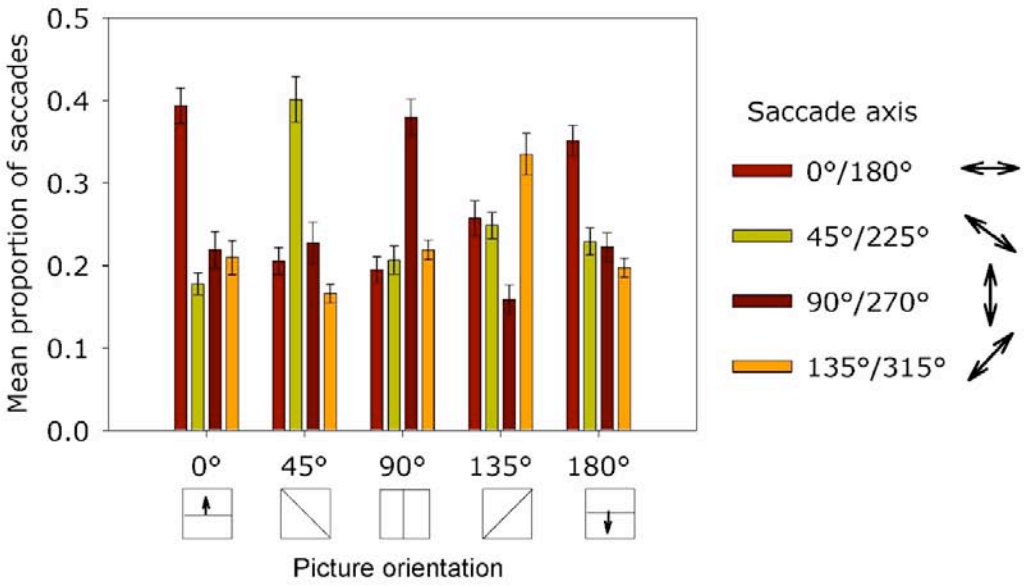


Figure 10.5. Mean proportion of saccades within each saccade axis, as a function of picture orientation. Error bars indicate plus/minus one standard error of the mean across participants.

Saccade direction over time

How early does the orientation of the image begin to effect the eye movement direction? To investigate this I looked at the frequency of saccades in each axis as a function of ordinal saccade number. As previously the first saccade was removed and to remain consistent the following saccades are numbered from 2 to 6. For statistical analysis I pooled data across the picture orientation conditions by rotating each saccade population so that the orientation of the picture was aligned with the horizontal. The four axes can then be thought of as relative to the dominant (horizon) axis. A two-way repeated-measures ANOVA was then possible with axis and ordinal saccade number (from 2nd to 6th) as factors. The data for this analysis is shown in Figure 10.6. Across the five saccades the orientation bias is shown by a main effect of axis, $F(3,36)=33.38$, $MSE=13.55$, $p<.001$. Pairwise comparisons showed a predominance of 0˚ saccades (versus 45˚, $t(12)=6.70$, $p<.001$; versus 90˚, $t(12)=5.84$, $p<.001$; versus 135˚, $t(12)=7.96$, $p<.001$). There were no other reliable differences. There was also an interaction showing that the asymmetry in saccade direction varied with ordinal saccade number, $F(12,144)=2.37$, $MSE=9.92$, $p<.01$. Planned comparisons showed that there were more

saccades made in the 0˚ axis than those in other directions (averaged across levels) and that this was the case for all saccades except the second (on 2nd saccade, $t(12)=1.47$; 3rd, $t(12)=5.13$, $p<.001$; 4th, $t(12)=3.51$, $p<.005$; 5th, $t(12)=4.52$, $p=.001$; 6th, $t(12)=5.78$, $p<.001$). Thus the pattern in saccade directions emerges relatively early on the second free saccade.

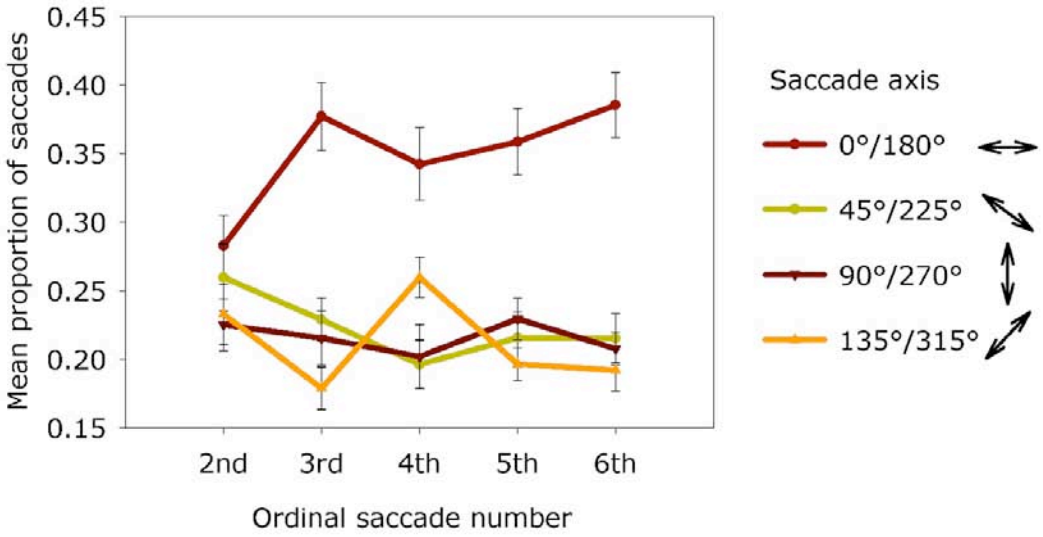


Figure 10.6. Mean proportion of saccades in each direction (with standard error bars) as a function of ordinal saccade number. Data is collapsed across orientation conditions, with saccade axis being relative to the original horizon.

Saccadic amplitude

A further question concerns the amplitude of saccades in each direction. The stimuli used here were square so did not require larger saccades in any particular direction. Are there also asymmetries in saccade length that vary according to picture orientation? Figure 10.7 shows that there are, with mean saccade amplitudes showing a similar pattern to that of the saccade directions in Figure 10.5. ANOVA showed a marginally significant main effect of picture orientation, $F(4,48)=2.47$, $MSE=1.54$, $p=.057$, and there was no effect of axis, $F(3,36)=2.29$, $MSE=2.26$, $p=.095$. There was an interaction of axis and picture orientation, $F(12,144)=8.59$, $MSE=1.81$, $p<.001$. Overall, saccades within the picture's original horizontal axis were larger than those within the other directions, although this effect was not as large as that seen with saccade frequency. As previously, planned comparisons quantified this, and in all cases saccades were longer in this dominant

orientation than the mean of the other directions (at 0˚ orientation, $t(12)=5.66$, $p<.001$; at 45˚ orientation, $t(12)=9.56$, $p<.001$; at 90˚ orientation, $t(12)=3.29$, $p<.01$; at 135˚ orientation, $t(12)=2.39$, $p<.05$; at 180˚ orientation, $t(12)=2.38$, $p<.05$). Looking at Figure 10.7, the trend is less clear when the picture was oriented at 90˚, and in this case vertical and horizontal saccades had a similar mean amplitude. The effect is largest in the 45˚ orientation, although this may be because, due to the square frame, an oblique horizon was in fact slightly longer.

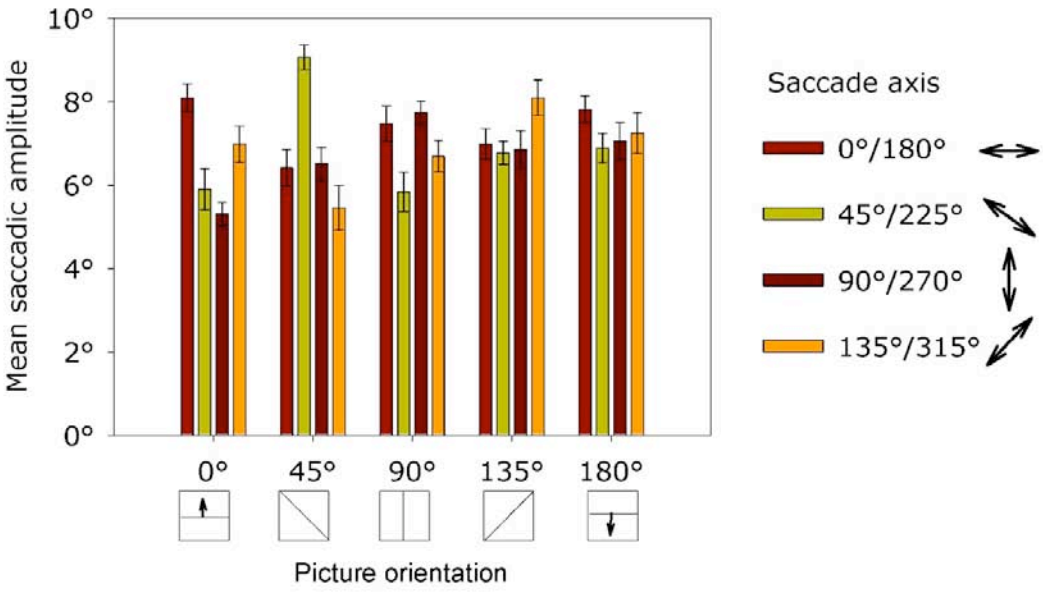


Figure 10.7. Mean saccadic amplitude (with error bars indicating the standard error of the mean) as a function of direction and picture orientation

10.2.3 Discussion

It is clear from these findings that there is a strong systematic tendency for saccades to occur along the axis of the natural horizon. As this tendency changes when the picture is rotated, an inflexible oculomotor account that fully explains the bias in terms of asymmetries in muscle control can be discounted. This does not contradict findings that horizontal saccades are faster or easier to make (Becker, 1991), and they may be more common in natural behaviour. However, people can suppress horizontal eye movements and make a larger proportion in other directions if the stimulus is oriented in a different way. The fact that the distribution of saccade directions changed on a trial-by-trial basis suggests that horizontal saccades are not just habitual patterns that have been learned and

cannot be altered. The square image frame and scattered starting positions means that a horizontal bias is not just due to the laboratory artefacts of a rectangular display (although the bias may act in conjunction with a default strategy of moving towards the centre of an image). There was also an interesting converging result in the data for saccadic amplitude, showing that people make saccades in the horizon axis that are on average longer than those made in other directions.

Two particularly interesting questions emerge from the findings. First, and particularly relevant to this thesis, could the pattern of saccades arise as a consequence of the distribution of saliency? It was especially true in the landscape stimuli used here that edges and other visual features tended to be clustered around the horizon. If these features attract eye movements, as suggested by saliency map models, then horizontal saccades are expected purely because that is where conspicuous information is located. This information might also have been the most useful to perceive and remember in order to perform the sentence verification task. However, an alternative explanation might rely on the preattentive recognition of gist or layout. By this account people perceive the orientation of a picture very quickly, and they use this with their knowledge of horizons to determine where information is likely to occur. One way of studying this is by asking when the bias for horizontal saccades occurs. It appears that the preference for saccades in the same axis as the horizon occurs early, although not immediately. It has been shown that gist information becomes available very early (Potter, 1976), and that it can influence initial eye movements (Castelhano & Henderson, 2007), although the current data did not find an effect on the first two saccades. It has also been argued that the influence of salience is greatest nearer the start of viewing a picture (Parkhurst et al., 2002). Experiment 10(b) compares two different types of scenes to see if the presence of a clear horizon can explain the scanning pattern.

A second consideration is whether the biases in viewing have a deleterious effect on processing of the scene. With this in mind participants in Experiment 10(b) were tested later to see how well they had encoded the rotated images. There has been considerable interest in whether the eye movements made when encoding an image affect those made when viewing it again (Althoff & Cohen, 1999; Noton & Stark, 1971;

Chapter 5). For this reason Experiment 10(b) explored whether there are any systematic effects of saccade direction at encoding on eye movements made whilst viewing the same images in a memory test. Do prior orientation and the resulting pattern of saccade directions have an impact on later re-viewing and recognition?

10.3 Experiment 10(b): scene type and orientation

In this experiment landscapes were compared with interior scenes. As interiors do not feature a natural horizon, and tend to have less clustering of features, a purely bottom-up account should lead to a reduced bias for saccades that follow the horizon. An explanation that relies only on early recognition of scene layout need not distinguish between landscapes and interiors and so would predict a similar bias in both cases.

In addition, participants are later given an unexpected memory test to see whether their recognition memory is affected by prior orientation and the resulting scanning strategy. This task also gives the opportunity of looking at saccade biases at recognition, in addition to those during an encoding task. Will the demands of this task alter the saccade bias?

10.3.1 Method

Participants

Twelve participants took part that had not been tested previously. All participants were from the University of British Columbia and took part for course credit. They matched the other participants tested in this thesis in all other respects.

Stimuli, apparatus and design

These were very similar to those used in Experiment 10(a). The same 40 landscapes were used here as in Experiment 10(a). In addition, the same number of interior photographs was also used. These were colour photographs at the same resolution as the landscapes, and they were prepared in the same way to give a selection of different rotations. Only half of the resulting 80 stimuli were presented with sentences in the first part of the task, whilst the other half were presented at the normal orientation in the recognition phase.

In order to describe the distribution of features in the two types of stimuli they were analysed using two different methods. Saliency maps were created for all stimuli.

An average saliency map was then created for the landscapes and the interiors, by simply adding the saliency (pixel intensity) values for each point to those for the same spatial location in all other images and then dividing by the number of images. The resulting average maps are shown in Figure 10.8(a), and it is apparent that the salient points in the landscapes tend to be near the horizon, whilst saliency is more distributed in interiors. To test this I divided the images into three horizontal bands and compared the average saliency in the central third with that in the top and bottom bands. Looking at the landscape stimuli, the mean saliency was higher in the central third of the picture than in the outer two thirds (paired *t* test across stimuli, $t(39)=4.4$, $p<.001$). However, in interiors saliency was not reliably greater in the centre than elsewhere ($t(39)=1.3$, $p=.18$).

If the horizon is a potent cue for saccade orientation, then horizontally oriented edges might be particularly important. Coppola et al. (1998) reported that natural scenes tend to have a predominance of oriented contours at the cardinal, rather than the oblique orientations. To see if this was the case in our stimuli, I analysed the edge content of the landscapes and interiors using the same method as Coppola et al. (1998). Images were converted to greyscale and analysed using a simple Sobel filter and the software MATLAB. Convolution with a three pixel square kernel gave the local gradient at each point in the image for horizontal and vertical directions. Combining these outputs resulted in an estimate of the orientation and magnitude of contours at each point in the image (see Coppola et al., 1998, and Appendix C for further details). Figure 10.8(b) plots the summed magnitude for each orientation: the frequency of each direction weighted by the magnitude of the gradient. The results, which are similar to those in Coppola et al. (1999), show that the distribution of edges is biased towards horizontal and vertical orientations, with relatively fewer oblique contours. It is interesting to note that while both types of picture contain many horizontal lines, interiors are more likely to have vertical edges. Specifically, the summed magnitude of near (±22.5°) horizontal orientations was greater than that in near vertical orientations in landscapes ($t(39)=5.0$, $p<.001$), but in interiors both horizontal and vertical orientations were just as strongly represented ($t(39)=1.7$, $p=.10$).

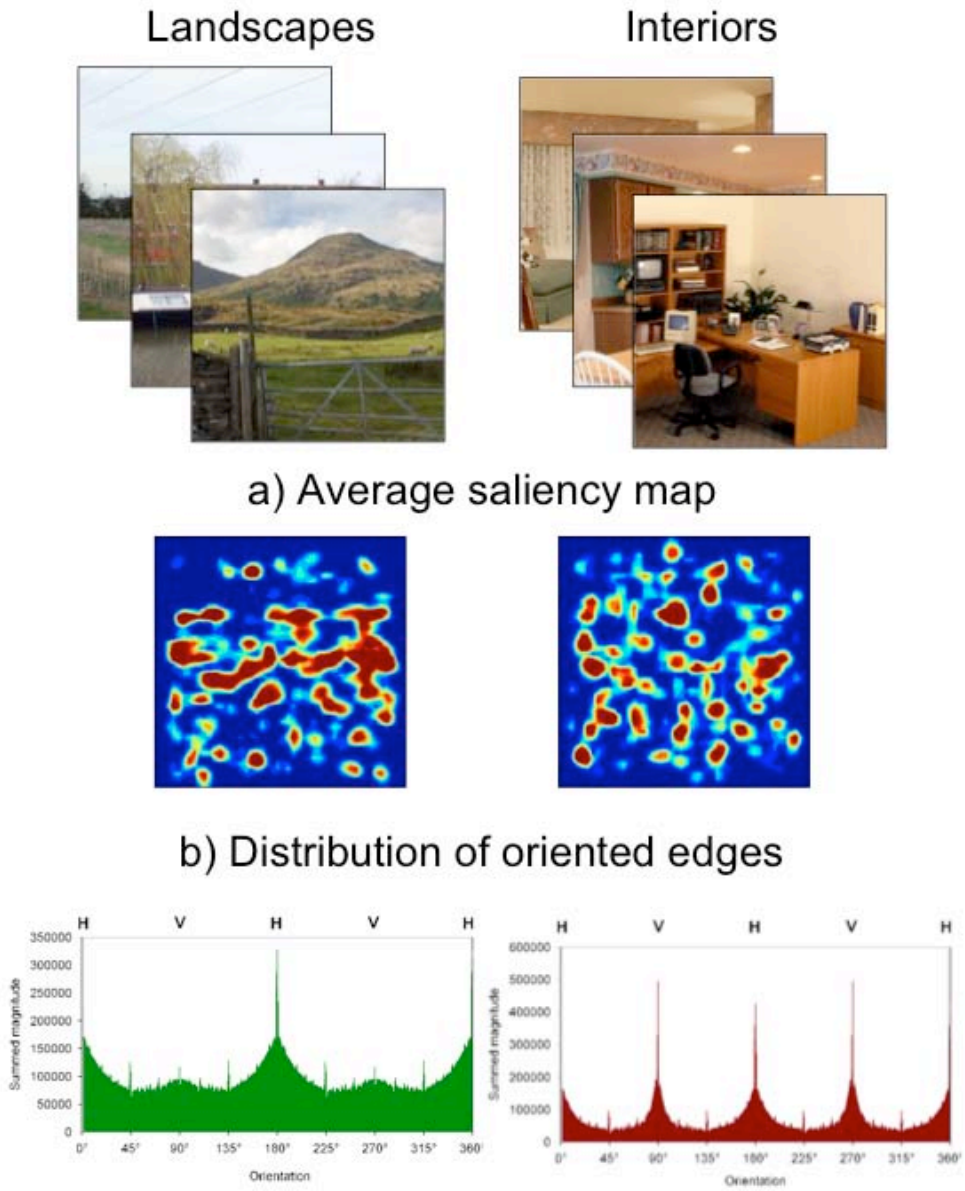


Figure 10.8. Analysis of features in the different types of stimuli. Maps show average saliency (a, with warmer areas indicating higher saliency) and the distribution of oriented contours (b). Plots in (b) show the mean summed magnitude across all the pictures in the set, reflecting the frequency and gradient intensity of contours at each orientation. The full range of orientations is shown, although symmetrical orientations (0/180, 90/270 etc) are equivalent and feature only because the edge filter distinguished between gradients from dark to light and those from light to dark. The images contained an abundance of horizontal (H) and vertical (V) edges.

Investigating the features of landscapes and interiors in this way allows one to make some general observations about the two types of stimuli. In landscapes, points that stand out from their surround tend to be clustered near the horizontal midline, and there are more horizontal than vertical edges and fewer still oblique contours. In interiors,

however, salient features are more distributed. There is no pronounced horizon and there are more vertical edges, presumably because of the presence of walls and the manmade edges of objects. If eye movements are following these features their direction should be differently distributed in the two types of scene.

The eye-tracking set-up was arranged as in the previous experiment. The same make and model of eye tracker was used (EyeLink II), and a validation procedure showed that it was recording at a high spatial resolution of at least 0.5˚.

Procedure

The first part of this experiment was the same as that in Experiment 10(a). The encoding stimuli were presented in a random order, and each one was followed by a sentence verification test. Landscapes and interiors were intermixed. This order was chosen as opposed to blocking by picture type, so as to avoid participants becoming accustomed to one type of picture and adopting a less dynamic strategy. As previously, participants were instructed to pay attention to the stimuli so that they could verify the sentence as accurately as possible.

Participants were then given an intervening task consisting of viewing fractal images, a task which took approximately fifteen minutes. Once this task was complete, a surprise memory test was presented with the images from the sentence verification test. Participants had no reason to suspect this would occur. Following a drift correct fixation marker in the centre of the screen, each test picture, half of which had been seen previously, was presented for 2500 ms. This duration was the same as used in the encoding phase, allowing a similar number of saccades to be compared. All trials in this part of the experiment began in the centre of the screen, unlike picture viewing in the encoding phase. This meant that (in)congruency in starting location between first and second viewing was equal across all orientations. All pictures were presented at the normal, non-rotated orientation. Following picture offset, a response screen asked subjects to respond with one of two keys whether the picture had been seen previously. One of the points of interest in this experiment was whether previous orientation affected responses at recognition, and so it was also useful to know whether participants remembered at which orientation pictures had been presented. With this in mind, if

participants responded that they had seen the picture before an additional screen asked them to indicate which way the picture had been rotated. There were five possible responses (0,45,90,135 and 180˚) and these were tied to five different keys on the keyboard.

10.3.2 Results

As in Experiment 10(a), the proportion of correct answers in the sentence task was relatively high (*M*=0.771) and these responses will not be considered further. The recognition test results, however, can tell us how well the pictures were processed in the general encoding-for-sentence-verification task. After looking at whether this encoding was affected by orientation the remaining analyses concentrate on whether a horizontal bias exists in both landscapes and interiors, and whether it changes when reviewing pictures for the second time.

Recognition test responses

How accurate were participants at recognising the previously rotated scenes? The mean proportion of hits (the proportion of images seen previously, at any orientation, which were recognised as such) is shown in the top half of Table 11.1, as a function of previous orientation and scene type. The false alarm rate was generally low (*M*=0.126, *SEM*=0.024). There was an effect of orientation on recognition, $F(4,44)=17.76$, $MSE=0.0372$, $p<.001$, but there was no effect of scene type, $F(1,11)<1$, and no interaction, $F(4,44)=1.82$, $MSE=0.0435$, $p=.141$. Both landscapes and interiors were recognised equally well and showed a similar pattern of hits across the different rotations. Comparing between orientations, recognition was higher for those pictures that had originally been presented normally (ie. at 0˚) when compared with all other levels (versus 45˚, $t(11)=7.70$, $p<.001$; versus 90˚, $t(11)=6.05$, $p=.001$; versus 135˚, $t(11)=3.85$, $p<.05$; versus 180˚, $t(11)=5.85$, $p=.001$). There were also some other reliable differences, namely that the 135˚ condition led to better performance than elsewhere (versus 45˚, $t(11)=4.65$, $p<.01$; versus 180˚, $t(11)=4.27$, $p=.01$). Chance performance in this task was 50%, and a one-sample *t* test compared each condition against this value. Whilst performance in the 0, 90 and 135˚ rotated condition was better than chance that in the 45

and 180˚ conditions was not (0˚, $t(11)=15.28$, $p<.001$; 45˚, $t(11)<1$; 90˚, $t(11)=2.56$, $p<.05$; 135˚, $t(11)=5.14$, $p<.001$; 180˚, $t(11)<1$).

			Original picture orientation				
			0	**45**	**90**	**135**	**180**
Proportion of hits	**Landscapes**	***M***	0.833	0.569	0.646	0.729	0.542
		SEM	*0.036*	*0.086*	*0.057*	*0.065*	*0.103*
	Interiors	***M***	0.958	0.396	0.583	0.785	0.563
		SEM	*0.028*	*0.084*	*0.071*	*0.061*	*0.045*
Proportion correctly recognised orientation	**Landscapes**	***M***	0.688	0.042	0.083	0.062	0.375
		SEM	*0.076*	*0.028*	*0.047*	*0.033*	*0.092*
	Interiors	***M***	0.667	0.069	0.083	0.146	0.125
		SEM	*0.094*	*0.048*	*0.036*	*0.048*	*0.038*

Table 10.1. Mean and standard errors for responses to the memory test in Experiment 10(b).

An additional question was whether participants could remember at which rotation a picture had been seen. It is clear from the mean proportion of correct orientation responses (Table 10.1 bottom half) that performance on this task was poor for all but the pictures shown at 0˚. This was seen in an effect of orientation on accuracy, $F(4,44)=28.09$, $MSE=0.057$, $p<.001$, such that 0˚ led to higher accuracy than the other orientations. Orientation recognition accuracy was similar for both picture types, $F(1,11)<1$. Comparing between orientations, the 0˚ condition led to much better performance than elsewhere (versus 45˚, $t(11)=6.70$, $p<.001$; versus 90˚, $t(11)=6.51$, $p<.001$; versus 135˚, $t(11)=6.98$, $p<.001$; versus 180˚, $t(11)=5.02$, $p<.005$). Of the remaining comparisons, only 180˚ (upside down) presentations were significantly different from the 90˚ condition ($t(11)=3.97$, $p<.05$). As there were five possible responses, a chance level of accuracy would be 20%, and only the 0˚ condition led to reliably better-than-chance performance ($t(11)=6.27$, $p<.001$). There was also an interaction between scene type and orientation, $F(4,44)=3.44$, $MSE=0.029$, $p=.016$, which was driven by a difference

between accuracy to 180˚ pictures. Accuracy was similar between landscapes and interiors for all orientations with the exception of the 180˚ condition, where there was a simple main effect of picture type, $F(1,11)=11.78$, $MSE=0.032$, $p<.01$. These upside-down pictures were remembered as such more accurately in landscapes than interiors and only above chance in the former. Thus, although all items at test were presented at their normal orientation, the orientation at which they had been presented previously was important.

Saccade direction at encoding

The first part of this experiment was concerned with whether the distribution of saccade directions was affected by the orientation of the picture, and in particular whether this effect was different in landscapes and interiors. As previously the first saccade was excluded as it was assumed that it was generally directed to the centre of the screen and that it was therefore dependant on the (randomised) starting location. The plots in Figure 10.9 show the proportion of saccades in each direction as a function of picture type. Figures 10.9(a) and 10.9(b) show the distribution of saccade directions within the four axes, as a function of picture orientation, and for landscapes and interiors respectively. These data are broadly similar to those in Figure 10.5. The variability is somewhat larger, probably due to a smaller number of experimental trials.

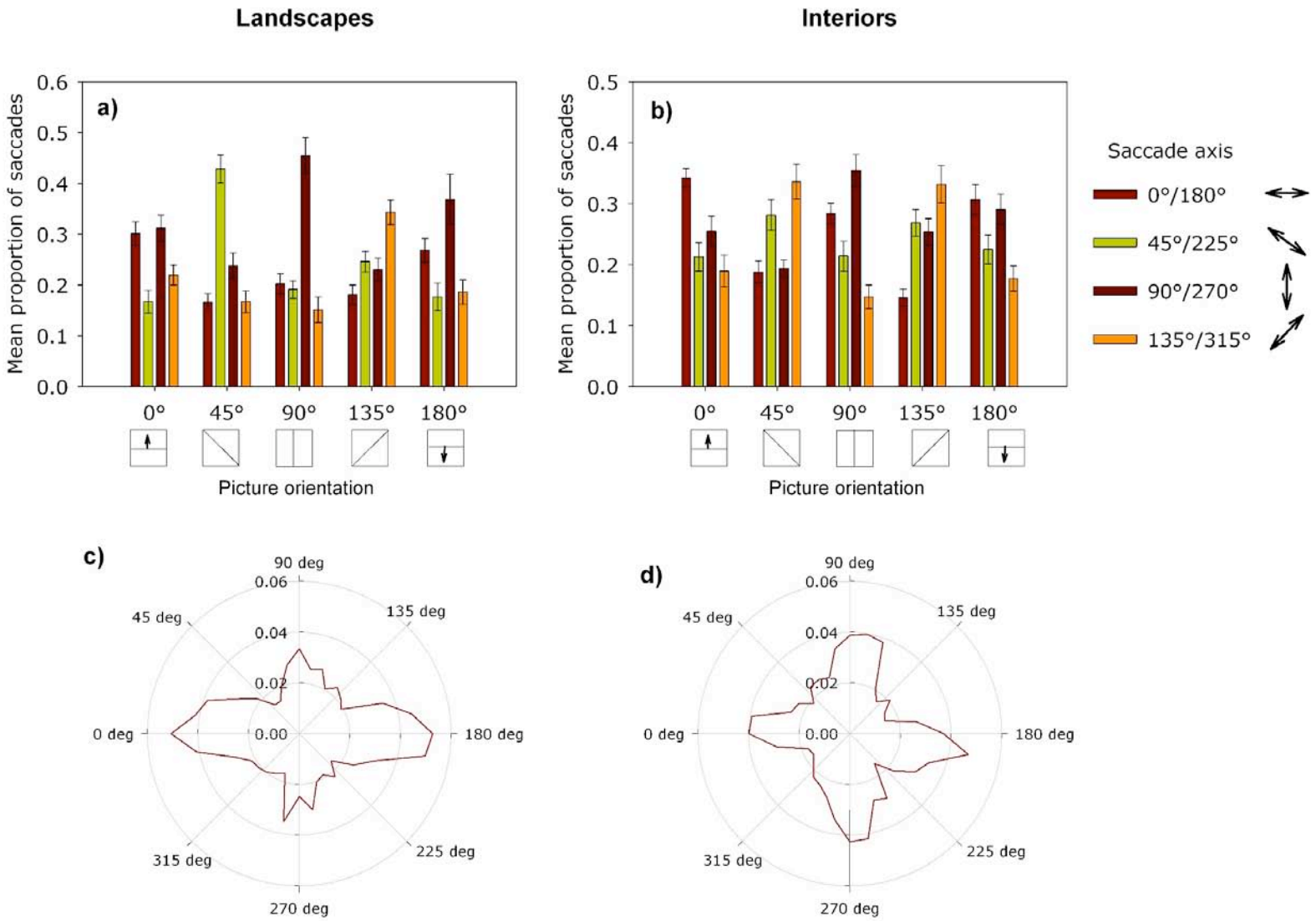


Figure 10.9. Mean proportion of saccades in each direction in landscapes (left panels) and interiors (right panels) during the encoding task. Panels a) and b) show the mean proportion of saccades in each axis as a function of picture orientation and with error bars indicating plus/minus one standard error across participants. In panels c) and d) data are shown for the full range of directions and collapsed across orientation conditions, with saccade axis being relative to the picture's original horizontal.

Two separate repeated-measures ANOVAs were computed to analyse the effects of picture orientation and saccade axis on saccade frequency in landscapes and interiors. As in Experiment 10(a) it was the interaction that was most important and I predicted that the most frequent direction would change with the orientation of the picture. In landscapes, there was no main effect of picture orientation, $F(4,44)=1.29$, $MSE=1.16$, $p=.29$, but a reliable effect of saccade axis, $F(3,33)=8.69$, $MSE=17.42$, $p<.001$. Regardless of picture orientation there were slightly more saccades in the vertical axis than in other directions (versus horizontal saccades, $t(11)=3.21$; versus 135˚, $t_{11}=3.43$, both $p<.05$). As predicted, the interaction was significant, $F(12, 132)=13.62$, $MSE=8.60$, $p<.001$, and planned t tests compared the frequency of saccades in the axis parallel to the horizon with the mean of the other three directions. This comparison was reliable in all picture conditions except 180˚ rotation. For instance, when the landscape's horizon was tilted at 45˚ there were more saccades in the 45/135˚ bin than elsewhere, and this was true for the direction parallel to the horizon in each of the first four orientation conditions (0˚ rotation, $t(11)=2.23$, $p<.05$; 45˚, $t(11)=7.58$, $p<.001$; 90˚, $t(11)=5.63$, $p<.001$; 135˚, $t(11)=3.74$, $p<.005$). In upside-down pictures this comparison was not reliable ($t(11)<1$). Whilst the trend in the 0 and 180˚ sets is not as clear-cut as in Experiment 10(a), due to a large number of vertical saccades, the overall pattern is the same: the dominant saccade direction shifts with the horizon.

Looking now at the interiors (Figure 10.9(b)) the change in distributions is qualitatively similar to that in landscapes. There was not a reliable main effect of picture orientation, $F(4,44)=2.26$, $MSE=1.15$, $p=.078$, or saccade axis, $F(3,33)<1$, but again there was an interaction, $F(12,132)=10.95$, $MSE=6.91$, $p<.001$. The predicted pattern—that the modal direction in each condition would be that parallel to the picture's original horizontal—was observed in all cases except the 45˚ rotation (where there were more saccades in the 135/315˚ bin). The planned comparisons confirmed this (0˚ rotation, $t(11)=6.25$, $p<.001$; 45˚, $t(11)=1.19$, $p=.26$; 90˚, $t(11)=4.15$, $p<.005$; 135˚, $t(11)=2.61$, $p<.05$; 180˚, $t(11)=2.20$, $p=0.05$).

To facilitate a comparison between the two types of scene Figures 10.9(c) and 10.9(d) plot the data from all orientation conditions, rotated so that the saccades counted as horizontal (0/180˚) are those made in the same direction as the picture's normal horizontal, and all other directions are measured relative to this. While the expected bias for horizontal saccades is clear in landscapes (Figure 10.9(c)), in interior pictures (Figure 10.9(d)) the shape of the plot is somewhat different. To characterise this statistically, I divided the arc into the four symmetrical axes (0,45,90 and 135) and looked at the frequency of saccades in each axis relative to the original orientation as a function of picture type. These data were analysed by a two-way repeated-measures ANOVA with axis (0,45,90 and 135˚ from the horizon) and picture type (landscape or interior) as factors. Although picture type was not reliable, $F(1,11)<1$, there was a main effect of relative axis, $F(3,33)=40.3$, $MSE=62.6$, $p<.001$. Across both types of picture there were more saccades in the axis of the original horizon (0˚) than elsewhere (planned *t* tests; versus 45˚, $t(11)=16.23$, $p<.001$; versus 90˚, $t(11)=3.33$, $p<0.05$; versus 135˚, $t(11)=17.00$, $p<.001$). There were also more 90˚ saccades than oblique angle saccades (versus 45˚, $t(11)=3.58$, $p<.05$; versus 135˚, $t(11)=4.23$, $p<.01$), though the frequency of saccades in the two different oblique axes did not differ.

Interestingly there was also an interaction, although this was only marginally significant, $F(3,33)=2.86$, $MSE=38.6$, $p=.052$. It is clear from the figures that this interaction is due to a larger proportion of "vertical" saccades (that is, those perpendicular to the original horizontal) whilst viewing interiors. Looking at the simple main effect of picture type, this was reliable for saccades in the 0˚ bin, $F(1,11)=12.05$, $MSE=17.43$, $p=.005$, and marginally so for 90˚ saccades, $F(1,11)=4.22$, $MSE=17.43$, $p=.065$. Interiors had fewer saccades in line with the original horizon, and more saccades perpendicular to it, than landscapes. The two types of scene did not differ in the frequency of oblique saccades: no simple main effect at 45˚, $F(1,11)=2.60$, $MSE=17.43$, $p=.13$; at 135˚, $F(1,11)<1$.

To summarise the saccade directions at encoding, the following conclusions can be drawn. First, the distribution of these directions remains broadly constant in the picture reference frame; as in Experiment 10(a), and in the majority of picture

orientations, the most common saccades were those in the plane of the original horizontal. Second, this effect was subtly different in interiors, which showed a relatively higher frequency of saccades in the picture's original vertical axis.

Saccade direction at test

How was saccade direction affected by the recognition task? In this part of the experiment all stimuli were presented at the normal orientation, so one would expect the standard bias for horizontal saccades as seen in the zero degree condition in Experiment 10(a). Saccade direction plots are shown in Figure 10.10 for old, previously seen pictures as a function of original orientation (a-e), and for novel, unseen pictures (f). These plots, and the subsequent analyses, include the first saccade as in this part of the experiment viewing began in the centre so direction was unconstrained. Any differences in the eye movements made whilst viewing pictures at test cannot be because of their orientation and must be due to recognising or reprocessing them in some way. There are some noticeable differences. Although most plots show a horizontal bias this is not as clear as in previous analyses and in those pictures previously shown at 180˚ there is large vertical bias. However, there appears to be no systematic effect of prior orientation; it is not the case that orientation at encoding carries over to effect viewing when presented the correct way round at test. To ascertain any non-specific effect of previous exposure, the saccade distribution associated with old pictures shown in the encoding phase was compared to that from novel items in the recognition test. Both types of items were equally likely to be interiors or landscapes.

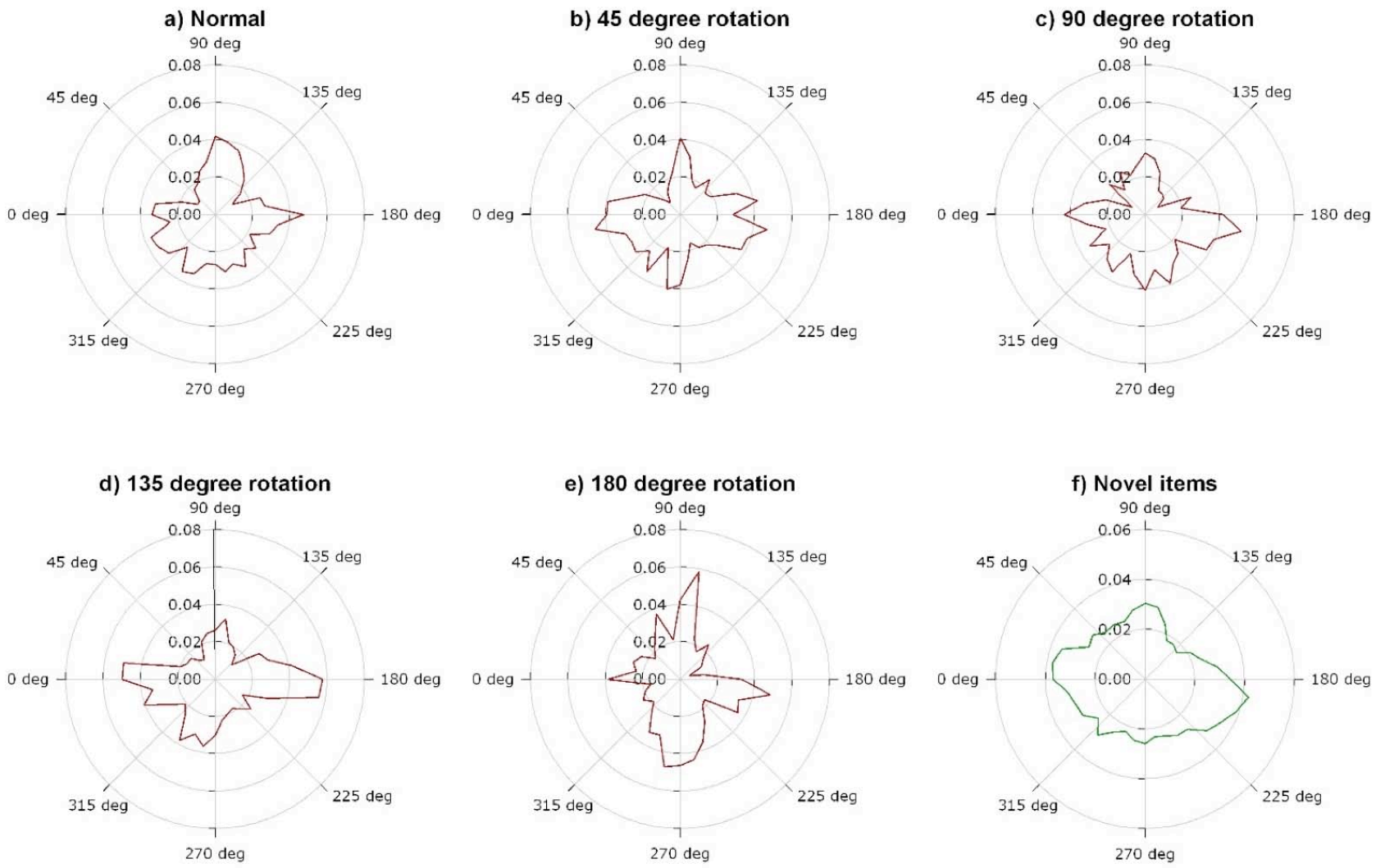


Figure 10.10. Saccade direction distributions for pictures viewed during the recognition test. All pictures were shown at the normal orientation, but some had previously been shown at various rotations (old pictures; a-e). New pictures had not been seen previously (f).

As previously the data were organised into the four major axes relative to the horizon and comparisons were performed on the frequency of saccades in each axis (Figure 10.11). The horizontal bias resulted in an effect of axis on saccade frequency, $F(3,33)=5.04$, $MSE=1164.5$, $p<.01$. Across old and new items more saccades were made in the horizontal than in the oblique axes (versus 45˚, $t(11)=4.79$, $p<.005$; versus 135˚, $t(11)=3.62$, $p<.05$). However, there was no difference between the frequency of horizontal and vertical saccades, and no other comparisons reached significance. There was no main effect of exposure on frequency, $F(1,11)<1$, but there was an interaction indicating that the distribution of saccades across different directions varied with exposure, $F(3,33)=19.52$, $MSE=52.0$, $p<.001$. It is clear from Figure 10.11 that the difference between the saccades made in old and new pictures is that there are more vertical saccades in the former. Looking at the simple main effects, exposure made a difference to the frequency of 90˚ saccades, $F(1,11)=51.7$, $MSE=33.2$, $p<.001$, and 45˚ saccades, $F(1,11)=38.4$, $MSE=33.2$, $p<.001$, but not those in the other two bins (both $F(1,11)<1$). Old items had more vertical saccades and fewer 45˚ saccades than new items at test.

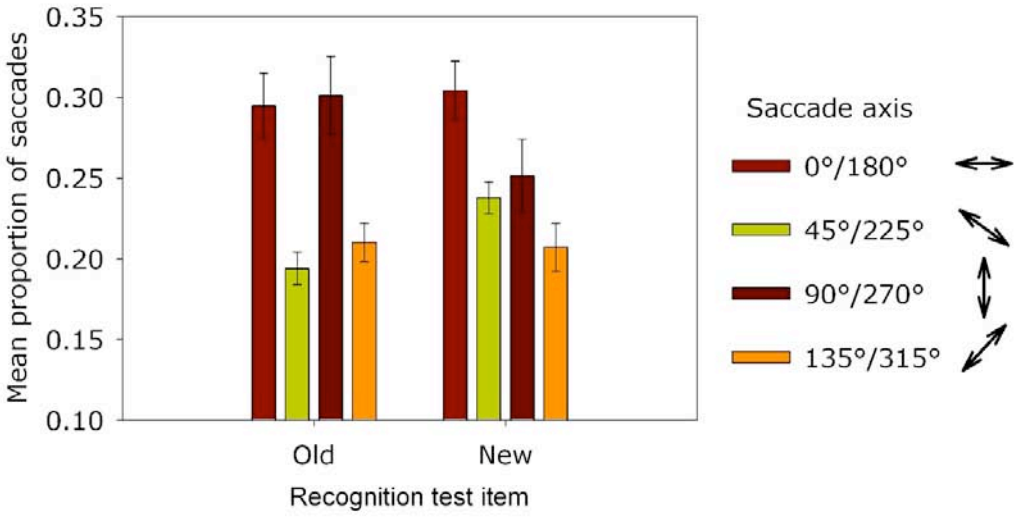


Figure 10.11. Mean proportion of saccades in the four axes, for old and new items when viewed in the recognition test.

Why did participants make more vertical saccades in previously seen images? If these saccades were associated with making a manual response (due to looking down at

the keypad for example) then they would also be seen in new trials, but they were not. However, it seems possible that they may have come about when observers recognised an image, perhaps because of the impending orientation recognition test. If the predominance of vertical saccades were due to this response then one would also expect it in cases where pictures were incorrectly recognised (false alarms). To test this, the mean proportion of all saccades that were in the vertical axis was compared between old trials that were correctly responded to (hits) and those that led to an error (false alarms; FAs). There was a significant difference ($t(11)=3.56$, $p<.005$). A greater proportion of saccades were made vertically when a hit occurred ($M=0.289$, $SEM=0.024$) than when an FA was made ($M=0.228$, $SEM=0.015$).

10.4 General Discussion

Experiments 10(a) and 10(b) measured several thousand saccades across people and pictures and found a robust bias for horizontal saccades. I am not aware of any other published results that explore saccade direction in natural images. When square pictures were presented at the normal orientation there were many more saccades made in the leftwards and rightwards directions than in the vertical or oblique directions. This supports previous reports, dating back to Brandt (1945). When the image was rotated, this pattern changed so that the most frequent direction for a saccade was parallel to the original orientation of the picture, suggesting sensitivity to the content of the image that can override the effects of rotation. Saccadic amplitude also varied systematically with direction (Experiment 10(a)); on average larger saccades were made in the axis of the horizon.

10.4.1 Explaining biases in saccade direction

With regard to the possible explanations for a horizontal bias from the introduction, several conclusions can be drawn. First, it is unlikely that the bias can be fully accounted for by laboratory artefacts. The predominance of horizontal saccades is not caused solely by rectangular stimuli. It would be interesting to look at saccades in pictures that are higher than they are wide (e.g. portraits), or in images with a circular frame, though the current findings suggest that horizontal saccades would still be more common. Even with

an earth-fixed frame of reference (a non-square monitor and the room behind) the degree to which the dominant eye movement direction rotated was highly systematic. Given the relatively difficult task there is no reason to think that participants were paying attention to cues outside the image, but the results suggest that if these cues were masked, removing any stable frame of reference, the effects of rotation would be even more pronounced. Outside the confines of a monitor, a fixed head position and a two-dimensional, discrete image, biases in saccade direction might change. Alternatively, in combination with head and trunk movements that tend to move horizontally, a horizontal saccade bias might be increased. This is a strong impetus for further research in more realistic settings. Although starting position was varied, the vast majority of trials began with a saccade into the centre of the image and thus it is possible that this central bias contributes to a predominance of horizontal saccades.

Both of the studies demonstrated that a strong oculomotor explanation should also be discounted. People can easily make saccades in the vertical and oblique axes if this is the way in which the picture is oriented. Similarly, although our environment may tend to be laid out horizontally, and our experience with this and other fixed situations (such as reading) may affect our propensity to move in any particular direction, observers can alter this within one or two self-initiated eye movements on a scene. There were no explicit instructions to alter scanning behaviour, and in Experiment 10(b) later memory for this orientation was poor, suggesting that it was not explicitly encoded.

So why do people move their eyes horizontally (or in parallel with the horizon of a rotated picture)? Two remaining possibilities from the introduction can be considered. The distribution of image features, such as edges or salient points, might guide attention in a bottom-up fashion. If these features were oriented or clustered along a horizontal axis of the picture then their distribution would change as the picture was rotated and this might account for the predominance of horizontal saccades. It may also have been that the task (verifying a sentence concerning objects and other details) biased participants to look at features that were distributed in this way. It has been shown elsewhere that fixation locations are dependent on expectations of target location in natural scenes (Neider & Zelinsky, 2006); modifying the task might change the importance of the

horizon and the frequency of horizontal saccades. In Experiment 10(b) interior photographs were contrasted with landscapes, to see whether the more distributed features in the former led to less of a horizontal bias. This was indeed the case, with the ratio of horizontal to vertical saccades changing from around 2:1 in landscapes to approach 1:1 in interiors. This is the pattern that would be predicted based on the analysis of edge content in the two types of picture, which showed relatively more vertical edges in interiors. Coppola et al. (1998) reported that cardinal orientations were over-represented in the natural world in comparison to oblique contours, and they linked this to the sensitivity shown by humans and other animals to different orientations. It is possible, therefore, that the reason people make fewer oblique saccades is due to the edges in the environment. Experiment 10(b) was not designed to manipulate edge content, but the results suggest an interesting link between contour magnitude and saccade direction. The landscapes used had more horizontal than vertical edges, and in most cases they led to a predominance of horizontal saccades. In interiors, where vertical edges were also prevalent, there was less of a horizontal bias and vertical saccades were common. This supports a bottom-up, image characteristics account of saccade direction control. At its strongest this account suggests that each subsequent saccade is targeted at the next most salient image feature (as in saliency map models), and that due to the distribution of saliency in landscapes these saccades tend to be moving between horizontally aligned points, whereas in interiors they are more spread out. The influence of edge information might suggest that the saliency model should weight the orientation channel more heavily.

A different explanation might posit the early recognition of scene type and layout. This fast perception of "gist" has been widely reported and can be predicted based on the analysis of low spatial frequency image statistics (Torralba, 2003). For example, if participants gleaned enough information from the first fixation to realise the type (landscape or interior) and orientation of the image, this information might activate stored representations of where objects and interesting features occur in this class of image. This information could then guide the eyes. Without a clear definition of the features that make up gist it is difficult to conclusively test their effect on eye movement

direction. Two other results from the present research are relevant for distinguishing between bottom-up control and early gist acquisition. First, when saccade direction was inspected as a function of time since picture onset, the disparity between the frequency of horizontal saccades and those made in other directions increased from the second saccade (where there was no significant difference) to the third saccade and beyond. Parkhurst et al. (2002) suggest that the influence of visual saliency declines over time, and if points on a saliency map are selected and then inhibited based on a winner-take-all system then to some degree later fixations should be made to less salient regions. Therefore, if horizontal saccades were due to the distribution of saliency, one would expect the bias to be greatest on the first free saccade and to decrease over multiple saccades, but this was not the case. It seems likely that gist and layout information build up over several fixations, so this could explain the increase in saccade bias following the first fixation on the picture. A second point of interest is the 180˚ rotation condition, where the image was completely inverted. Some researchers have reported that inverting a scene disrupts the acquisition of gist, so one might expect a different pattern of eye movements, even though the distribution of features relative to the horizontal axis will be the same as in the normally oriented picture. In fact, saccades in the 180˚ condition showed an equally strong horizontal bias.

There were additional noteworthy findings. In Experiment 10(a), saccades in the axis of the horizon also had, on average, larger amplitude. Due to the square dimensions of the image, oblique directions had a longer plane than cardinal directions in which to move, and so I will be cautious about the effect this may have had on some rotations. However, even in the 0, 90 and 180˚ conditions larger saccades were made in the plane of the horizon, despite the presence of a longer oblique axis running in a different direction. Saccadic amplitude could be taken as an indication of the degree of peripheral processing; greater processing of peripheral regions allow more distant saccade targets to be selected, leading to larger saccades. If this is the case then it suggests an asymmetry in the way processing of information away from fixation takes place, perhaps with covert attention spreading further along the perceived horizon than in the direction perpendicular to it.

10.4.2 Saccade direction and recognition

Experiment 10(b) looked at participants' later, incidental memory for pictures that had been rotated. Subsequent recognition was much better for pictures that had been shown the correct way up. Thus, even though people could adjust their scanning patterns in line with the rotation of the picture encoding may have been better when not rotated. Recognition of a picture's previous orientation was rather poor for almost all the conditions. This suggests that although the pictures were encoded into memory (as shown by above-chance old/new recognition), orientation was not remembered. Thus if an oriented representation of the scene is formed early in viewing to guide saccade direction, this representation is not maintained and available for later retrieval. The memory results were consistent for both landscapes and interiors. While these results are interesting they should be treated with some caution. Investigating memory was not one of the main aims of the work presented here, and there are several issues that might have affected the results. First, the cropping of rotated images meant that some of the correctly oriented test images contained slightly different information from when they were presented initially. The differences were small and peripheral, but it is possible that they may have led to more errors for rotated pictures. Second, further research is needed to unravel the effects of orientation at encoding and at test. In Experiment 10(b), the recognition advantage for non-rotated images might be due either to better encoding, or to the congruency between encoding and test orientation. The starting fixation location used at encoding and test, and the congruency between these, was controlled in the present study but this might also have an effect on memory. I am pursuing these memory effects elsewhere, but the remainder of the discussion will concentrate on the eye movement data.

How did re-exposure to pictures in the test phase of a memory test effect scanning? Scanpath theory suggests that eye movement sequences are stored along with the features of an image, and recapitulated when that image is seen again (Noton & Stark, 1971). If this were the case then one might expect carry-over effects of the dominant scanning direction on the eye movements made at test. However, there was no systematic effect of prior orientation on the direction of saccades made when pictures were viewed

for the second time. Although the present chapter does not look at scanpaths (chains of multiple, sequential saccades) this finding, along with others in the literature, suggests that the predictions of scanpath theory are too strong (Foulsham & Underwood, 2008; see also Henderson, 2003). Recent work by Althoff and Cohen (1999) has examined a "reprocessing effect" for pictures of scenes and faces, whereby there are differences in the eye movements made when a stimulus has been seen before compared with when it is novel, even in the absence of explicit recognition. The analysis of saccade direction in old versus new pictures at test gives an indication that a similar effect is happening here. Old pictures could be distinguished from those that had not been seen on the basis of the proportion of vertical saccades. Correct recognition of old pictures was associated with more vertical saccades, and this was not found in trials leading to a false alarm. It is unclear what caused this difference, although it may have been related to the requirement to make further orientation recognition responses. This observation also suggests that a top-down, task-driven factor can modify and reduce the tendency to make horizontal saccades.

10.4.3 Implications for models of eye guidance

What are the implications of the saccade pattern discussed for models that aim to predict eye movements in natural scenes? Bottom-up models, such as the saliency map model of Itti and Koch (2000), have been adapted to take into account the space-variant sampling of the retina (Vincent et al., 2007). However, locations situated in the horizontal, vertical and oblique directions are equally likely to become saccade targets in this model. The data presented here suggest that this is not the case, and that across a range of images horizontal saccades are more likely, and that oblique movements are rare. Saliency-based models might produce better predictions if they incorporated a saccade generator that took into account what is known about saccade dynamics and direction distributions. Of course, if the distribution of salient features is asymmetric then this pattern might emerge naturally, and the comparison with interior scenes, which have more distributed features and show less of a horizontal bias, supports this account. On the other hand several findings in the present research suggest that this bias varies according to gist and top-down goals – it forms after several fixations and is affected by previous viewings.

Perhaps a more realistic framework can therefore be provided by the contextual guidance model of Torralba et al. (2006). In this model local saliency is computed in parallel with the extraction of global features which can provide gist and layout information. This information provides contextual priors to bias the saliency map to certain locations, and if it included rough knowledge of scene orientation and the likely location of important features this would produce saccade asymmetries. The model was designed to predict real-world visual search for a known target object, so it would need to be generalised to account for the encoding task used here. I will discuss such a model in my conclusions in Chapter 12.

10.5 Conclusions and links forward

This chapter described a novel way of exploring eye movements in natural scenes by looking at the distribution of saccade directions. The bias for saccades parallel to the horizon is robust even in square photographs and yet it can be quickly adjusted if the picture is rotated. Models of eye movements need to be able to predict this observation based on either bottom-up feature distributions or, as seems likely, in concert with higher level knowledge of scene layout. In terms of the saliency map model, the robust horizon bias suggests that the model could be improved by favouring horizontal shifts, or perhaps by preferentially weighting the orientation channel.

11 Saliency and fixation patterns in visual agnosia

11.1 Introduction

In previous chapters I have suggested that bottom-up saliency and top-down knowledge combine or compete to guide the eyes, particularly when people are searching for a target. How is the link between saliency and fixation during scene perception affected by a disruption in top-down object recognition? This chapter addresses this question by looking at the behaviour of a patient with visual agnosia on some of the tasks previously discussed.

Visual agnosia is a neuropsychological impairment in which the patient is unable to recognise objects by sight, despite normal visual acuity and semantic knowledge (Farah, 1990; Humphreys & Riddoch, 1987; Riddoch & Humphreys, 1987). A good deal of research has described this impairment and its key features. In particular, researchers have attempted to distinguish between those whose impairment can be attributed to a relatively low level difficulty in perceiving shape and form (apperceptive agnosics) and those whose perception is intact but disconnected from its semantic associates (associative agnosics). Riddoch and Humphreys (1987) identified a third sub-type in their patient HJA, integrative agnosia, as a failure to combine stimulus features or attributes into a coherent object (see Behrmann, 2003 for a review of this and other cases). HJA showed poor visual object recognition but was able to copy line drawings he could not recognise and had detailed knowledge of objects that he could produce in response to verbal labels. He was particularly impaired at naming overlapping or occluded objects (Giersch, Humphreys, Boucart, & Kovacs, 2000), consistent with a difficulty in appropriately combining local form information. HJA's responses in visual search were disproportionately prolonged when the task involved integration of features across the display (finding an inverted T amongst upright T's) but not when targets were identified by a single feature (an oriented line amongst vertical lines; see also Delvenne, Seron, Coyette, & Rossion, 2004). Riddoch and Humphreys (1987) also reported that HJA's horizontal and vertical eye movements were normal.

The research on visual agnosia raises some interesting questions about how these patients inspect a natural scene, particularly when searching for a target object. There is some evidence that additional information such as colour improves the recognition performance of agnosics. For example, they may be better at recognising real objects and photographs than line drawings (Behrmann, 2003). HJA is also able to use contextual scene information to improve his recognition of objects (Riddoch & Humphreys, 1987). For these reasons it might be that agnosics are less impaired at searching in realistic scenes than at identifying line drawings.

An alternative prediction is that the prominence of bottom-up versus top-down factors in influencing eye movements might change in visual agnosia. Assuming that early perceptual mechanisms are undamaged, local stimulus features should affect the eye movements of agnosics but they might have difficulty linking these to what they know about the target or combining elements across the scene. In the search experiments in this thesis the semantic knowledge of objects had an impact on where fixations were distributed (they tended to be on targets, objects that were similar to targets, and places where targets were likely to be).

What happens when the ability to recognize target features or global scene properties is absent or impaired? Under such circumstances it might be more difficult to implement top-down guidance. A person with visual agnosia might not be able to override a bottom-up saliency-based system, if such exists, due their inability to link raw visual input to top-down knowledge. If so, their fixation patterns should conform more closely to the predictions of the saliency map model, and show a significant impact of the salience of stimulus properties on visual search in situations where normal subjects do not display such effects. Such a finding would extend the applicability of the saliency map model to more naturalistic situations and support the suggestion that such bottom-up effects are present but normally over-ridden by top-down considerations. To explore these ideas, in Experiments 11(a) and (b) I repeated the first two experiments from this thesis with a patient with visual agnosia.

11.2 Experiment 11(a): category search in natural scenes

In this section I report on the fixation patterns made during visual search by a patient with general visual agnosia. Despite reasonable visual acuity and peripheral fields, indicating functioning low-level vision, she is unable to recognize even simple three-dimensional abstract forms, or line drawings of common objects. Neuropsychological studies in eye movement research have been carried out in patients with neglect or parietal lesions (Barton, Behrmann, & Black, 1998; Shimozaki et al., 2003); visual field loss (Martin, Riley, Kelly, Hayhoe, & Huxlin, 2007); and simultagnosia (Clavagnier, Berger, Klockgether, Moskau, & Karnath, 2006; Rizzo & Hurtig, 1987). There has also been some interest in eye movements and face processing in prosopagnosia (Barton, Radcliffe, Cherkasova, & Edelman, 2007). However, no previous research has explored the eye movements of patients with visual agnosia in a natural search task.

Given her inability to identify objects, how would this patient distribute fixations in the stimuli from Experiment 1 when given the instruction to search for a certain category of object? In Chapter 3 I reported that participants could quickly fixate the target and that saliency did not make much of a difference. Here, I hypothesized that with an agnosic patient there should be less top-down guidance in this task than with normal controls. If visual saliency is computed earlier than, or independent from, object recognition then saliency would be predicted to have an effect on eye movements. Furthermore, with reduced top-down biases, the eye movements produced might be closer to a raw saliency map than those made by normal controls.

11.2.1 Method

Case description

CH is a 63 year-old right-handed woman with slowly progressive visual difficulties over a period of six years. She first noted trouble with reading, especially large type, although she could still write. Subsequently she had difficulty recognizing the faces of her friends, relying on their voices instead. She had problems locating household objects, for example in the refrigerator or on the kitchen counter, and often misreached for items like light switches and cups. She confused navy with black but otherwise believed her colour vision

to be normal. She later developed more problems navigating in familiar surroundings and required an escort on her visits to the clinic.

Her visual acuity was good (20/25 for single letters) and her visual fields were full, confirmed by Goldmann perimetry. Saccades and pursuit eye movements were normal.

Neuropsychological evaluation showed normal general knowledge, expressive vocabulary and comprehension. She was able to write to dictation, with occasional spelling errors, but was unable to read even single words, proceeded by deciphering one letter at a time. Digit span was 6 forwards and 4 backwards, and she performed in the average range on all tests of auditory verbal learning and memory. Abstract verbal reasoning abilities were average to superior.

Visual tasks were severely impaired, with difficulty on line bisection, cancellation and search tasks. Her object recognition was slow and required a lot of effort. She was severely impaired on the Boston Naming Test. When naming line drawings she frequently failed or misidentified objects, often focussing on small details (key = "see the O", chair = "grass…something to sit on"). She had difficulty naming three-dimensional shapes (cylinder = "cone", cube = "all these angles, octagon", pyramid = "triangles in it, a tent"), although she could name two-dimensional figures like squares and circles. She interpreted the Cookie theft picture as "a woman washing dishes, something spilling, it must be a restaurant".

Further perceptual tasks confirmed severe visual problems. Benton line orientation test showed 20% accuracy, in the severely deficient range, but curvature discrimination was normal. Her ability to judge the spatial configuration of dot patterns was impaired, scoring 56% correct with 2 dots and 22% correct with 4 dots (chance = 33% correct). Her ability to judge whether a triangle was symmetric or not was at chance (56% correct), a task controls do with 100% accuracy. Her ability to distinguish famous faces from anonymous ones was poor (d' = 0.12, versus d' for controls >2.5), though her imagery for famous faces was quite good, scoring 13/16 for facial features and 11/16 on overall face shape (Barton & Cherkasova, 2003).

A cerebral perfusion scan with technetium injection at 4 years after onset showed marked hypoperfusion of the posterior parietal and temporal lobes. CT scan 6 years after onset of symptoms revealed some general sulcal prominence with occipital predominance of enlargement of the lateral ventricles, consistent with a diagnosis of posterior cortical atrophy (Benson, Davis, & Snyder, 1988).

Control participants

I have already described a group of student participants taking part in a category search (Experiment 1(a), Chapter 3). These participants ranged in age from 19 to 30 years old and hence I will refer to these as young controls (YCs). Age is known to have an impact on visual search and oculomotor behaviour (Rabbitt, 1965; Scialfa, Thomas, & Joffe, 1994) and so I also tested a group of 10 age-matched controls (AMCs; 5 females). These participants ranged in age from 58 to 76 years old, with a mean age of 65, making them comparable with patient CH. All reported themselves as healthy, had normal or corrected-to-normal vision (all but 3 wore glasses) and had visited an optician in the last 3 years. A short questionnaire confirmed that none of these participants reported difficulties in finding or identifying objects around the home or in recognising the faces of friends or family, and that they had no other history of visual problems.

Stimuli, apparatus and design

The EyeLink 1000 eye tracker was used to record eye movements with CH. This system uses a desktop-mounted camera to track the pupil image and corneal reflection, with a sampling rate of 1000 Hz. As such it is very similar to the systems used elsewhere in this thesis, but it is less restrictive and so easier and more comfortable for use with a patient. The patient's head was placed in a chin-rest and head frame to minimize head movements, viewing a screen placed 60 cm from the corneal surface, with dim background lighting. AMCs were tested using the standard head-mounted EyeLink II system described elsewhere.

The stimuli for the category search were described in Chapter 3. There were 48 scenes, half containing a piece of fruit that could be medium or low saliency. For a closer

analysis of the agnosic fixations I also generated saliency maps for all stimuli in order to analyse the saliency at fixation.

Procedure

CH performed calibration well, although she sometimes found it difficult to find and focus on the fixation dot without an experimenter pointing to it on the screen. The subsequent procedure was exactly as that in the category search condition in Experiment 1, with participants attempting to answer the question "Is there a piece of fruit in the scene?" as quickly as possible. It was stressed that the participant did not need to identify the fruit and that not all pictures would contain the target. Before testing, the task was explained and CH demonstrated that she had intact knowledge of different types of fruit. The response was a verbal "yes" or "no" that the investigator recorded by pressing one of two keys on a keyboard. This made it easier for the patient and the AMCs to accomplish the task and reduced the likelihood of data loss from participants struggling to respond themselves. The key press terminated the trial.

11.2.2 Results

I first compared CH's overall performance in terms of proportion correct and reaction time. The eye movement analyses can be divided into three categories: 1) General observations about the number of fixations, their duration, and the amplitude of saccades during each trial; 2) Fixation behaviour in relation to the target piece of fruit and the most salient region in the scene; and 3) an analysis of the underlying saliency map value at fixation. If CH's fixation patterns are guided more by bottom-up effects of saliency than by top-down effects from target knowledge, then she should fixate the target less often, show effects of target saliency, and require more fixations before eventually fixating upon the target than controls. If the relationship linking fixation and saliency differs between the patient and controls then one might also expect her to fixate the most salient region more often, and for saliency at CH's fixation locations to be higher, on average.

In each case the agnosic data can be described in relation to the distribution of values provided by the control groups, and quantified by a *z* score. The degree to which age had an impact on the results is also of interest, so the two control groups were

compared using independent-samples *t* tests, or with group as a between-subjects factor in a mixed ANOVA with target saliency. Where possible, contrasts between low- and medium-saliency trials within the single subject CH were made using the chi-squared test (for nominal data).

Behavioural data

CH's responded correctly on 63% of trials, which is significantly below the accuracy of control subjects (YCs=92%, $z = 4.4$; AMCs=95%, $z = 6.9$; both $p<.001$; Figure 11.1, top). While she detected both low saliency targets (8 hits out of 12) and medium saliency targets (9/12), this was offset by a relatively high false alarm rate (11/24). Furthermore, although she was not required to name target or other objects, she did so on several occasions, but usually with a misidentification (e.g. she called a brightly coloured book "a watermelon").

Across all trials, CH's mean reaction time was 9983 ms (*SD*=5426 ms). Looking at correct target-present trials only, CH's mean RT was 9830 ms (*SD*=5845 ms) for low saliency trials and 7151 ms (*SD*=2473 ms) for medium saliency trials. These are extremely prolonged compared to both YCs (low, $z=34.3$; medium, $z=29.2$) and AMCs (low, $z=13.6$; medium, $z=13.2$; all ps$<.001$; Figure 11.1, bottom). Control groups were compared with a 2 (group, YCs vs. AMCs) x 2 (low vs. medium saliency) mixed factorial ANOVA. There was a large effect of group, $F(1,23)=27.7$, $MSE=225177$, $p<.001$, indicating that the AMCs took longer to respond on average. There was also a within-subjects effect of saliency, $F(1,23)=19.8$, $MSE=27052$, $p<.001$, though this interacted with group, $F(1,23)=6.8$, $MSE=27052$, $p<.05$. Thus target saliency did not affect RT in younger controls (post hoc $t(14)=1.5$, $p=.15$), but it did have a reliable effect in the AMCs ($t(9)=4.2$, $p<.005$), where medium-saliency objects led to faster RTs than low-saliency objects. CH also responded more quickly on medium-saliency trials than low-saliency trials, but there were too few data points (due to her poor accuracy) so this was not analysed further.

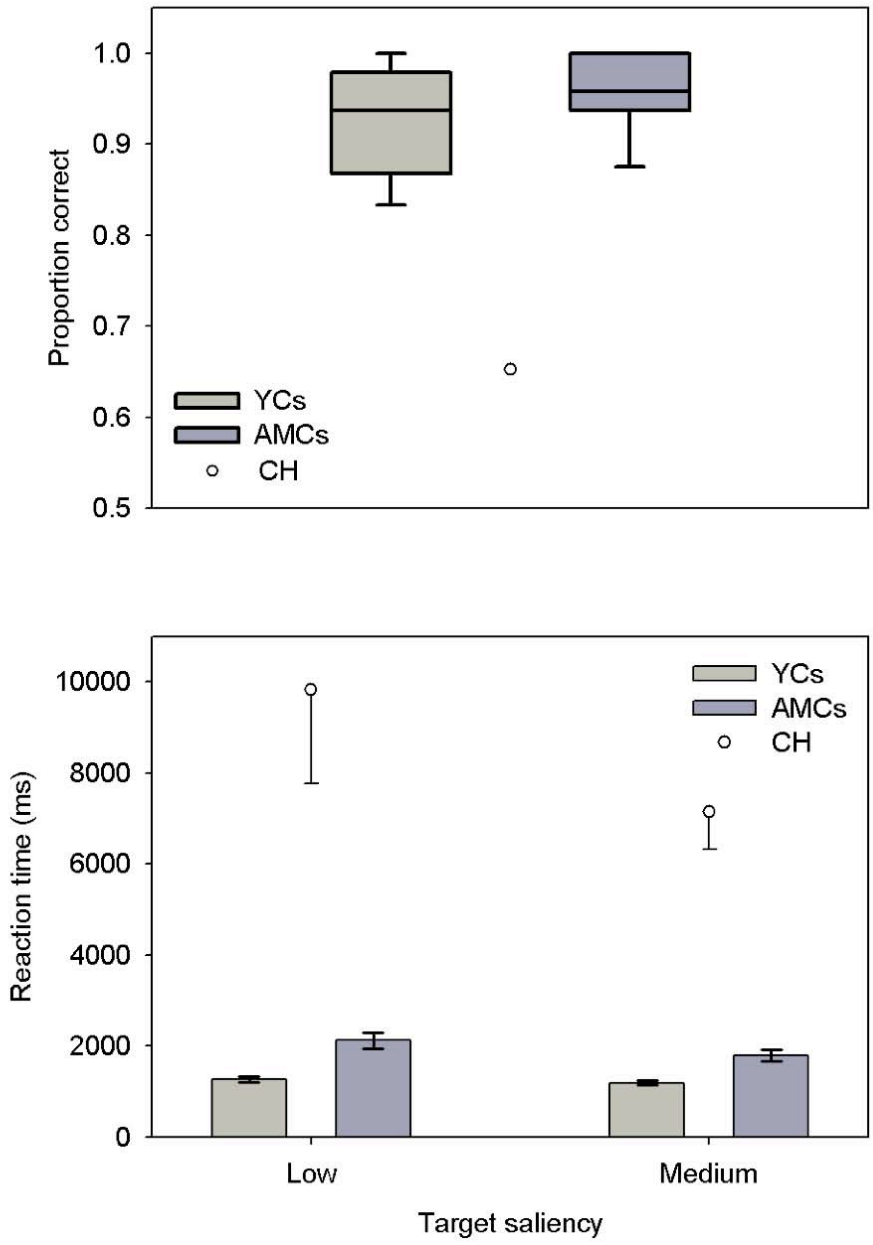


Figure 11.1. Proportion correct (top) and correct, target-present reaction time (bottom) for CH and the control groups. The overall proportion of correct responses is shown for controls as a box plot with the median and the 25th and 75th percentile (whiskers show 95th and 5th percentiles). Elsewhere, error bars show one standard error of the mean.

General oculomotor statistics

Figure 11.2 shows an example of fixation patterns for CH and control subjects, and these make it clear that her behaviour was quite different from normal observers. In the first example (top left panel), she fixates several salient regions but neglects the target. In the second scene (top right panel), she makes a large number of saccades and fixations and dwells on colourful and salient regions such as the lamp, completely missing the target. Control participants (bottom panels) make most fixations on the target, occasionally fixating other objects and avoiding blank areas and surfaces. The subsequent analyses aimed to quantify these observations across all trials.

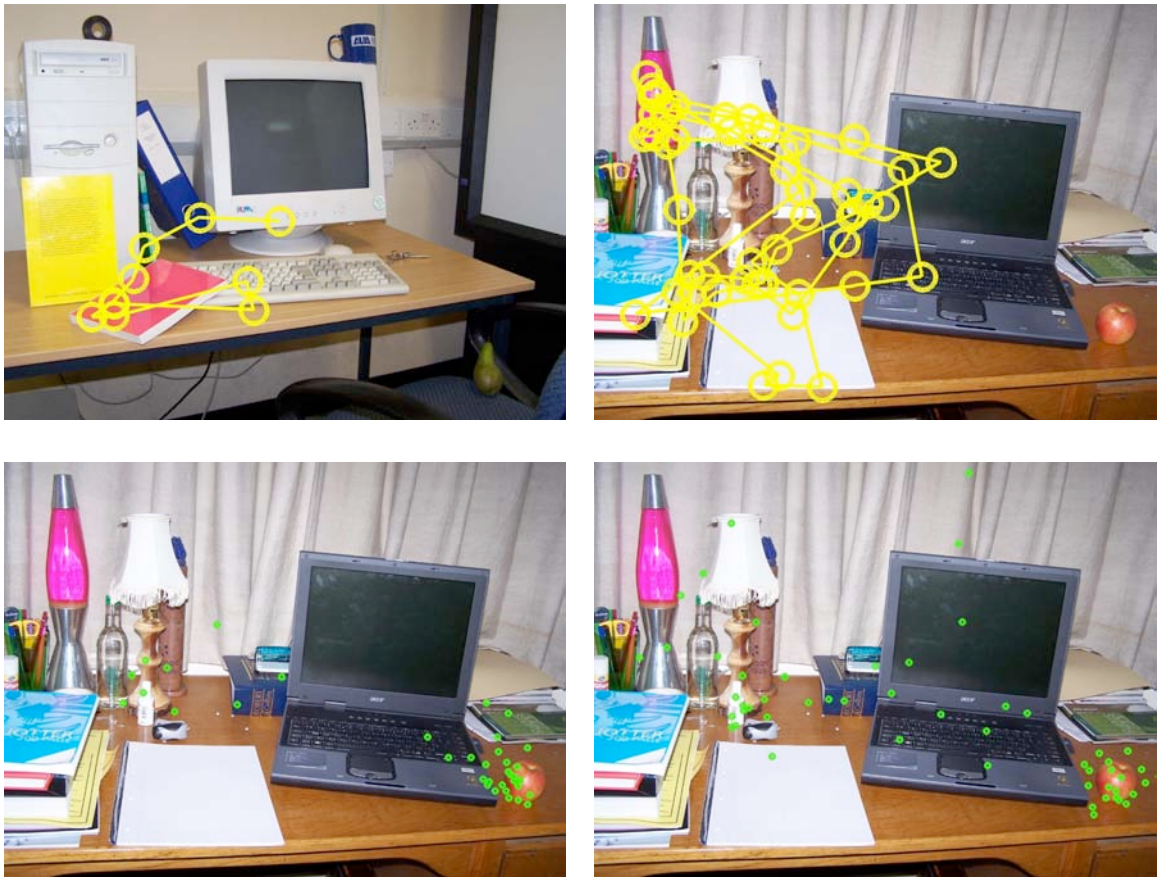


Figure 11.2. Examples of the fixation behaviour of CH (top row) and controls (bottom row). CH's fixations from a single trial are represented by yellow circles linked by lines showing saccades, and these are shown for the same stimulus as that in Figure 3.1 (top left) and for another scene (top right). The fixation locations of all control participants when looking at this scene are shown below, for YCs (bottom left) and AMCs (bottom right).

As the display was terminated when the target was found, the number of fixations per trial reflects the efficiency of visual search. Across all trials, CH made far more fixations (M = 36.7) than either YCs (M=6.6, z=14.4) or AMCs (M=12.3, z=4.6, both ps<.001). Table 11.1 shows the mean number of fixations in low and medium saliency trials and in target absent trials. The control groups were compared in a mixed factorial ANOVA with trial type as the within-subjects factor. As with RT the control groups were reliably different, F(1,23)=16.6, MSE=27.2, p<.001; AMCs made more fixations. Trial type also had an effect, F(2,46)=36.9, MSE=7.7, p<.001, with more fixations occurring in target absent trials than either low (t(24)=5.0) or medium saliency (t(24)=5.6) target present trials (both ps<.0001). There was also a reliable difference between low and medium saliency trials with fewer fixations in trials with a medium saliency target (t(24)=5.6, p<.001). The interaction was also significant, F(2,46)=5.07, MSE=7.7, p=.01; the AMCs and YCs were significantly different in all trial types (all at least p<.03), but this was more pronounced in the target absent trials.

The mean duration was calculated across all fixations, with the exception of the first, last and any that were on the target, so as to exclude processing time associated with responding. CH spent longer on each fixation than controls, although this was only reliable when compared with YCs (overall means, CH=230.2 ms; YCs=183.8 ms; z=1.79, p<.05; AMCs=205.0, z=0.83, p=.20). Looking at the different trial types, CH made longer fixations than the YC group in all cases, but this was not significant for target-absent trials (z=0.64, p=.26). The mean fixation duration of the two control groups was only marginally different (t(23)=1.9, p=.075).

CH also made smaller saccades (4.5˚ on average across trial types) than control subjects (YCs, M=9.1˚, z=3.16; AMCs, M=9.5˚, z=3.88, both ps<.001). The two control groups were not different (t(23)<1). Thus, to summarise the measures of global eye movement behaviour, CH made smaller saccades, more fixations and longer fixations whilst searching the scene. There were also indications that age had an effect; the older control group made more and slightly longer fixations.

		CH			YCs			AMCs		
		Low	**Medium**	**Target absent**	**Low**	**Medium**	**Target absent**	**Low**	**Medium**	**Target absent**
Number of fixations per trial	***M***	31.75	32.33	41.38	5.06	4.43	8.45	8.91	7.68	16.38
	SD	*19.30*	*19.20*	*22.20*	*1.00*	*0.60*	*3.80*	*3.16*	*2.71*	*8.18*
Saccadic amplitude (°)	***M***	4.07	4.75	4.66	8.59	8.49	10.27	9.12	9.18	9.92
	SD	*3.9*	*9.5*	*7.7*	*1.3*	*1.6*	*2.0*	*1.3*	*1.4*	*1.4*
Fixation duration (ms)	***M***	254	247	210	176	173	193	208	206	203
	SD	*56*	*56*	*25*	*28*	*31*	*27*	*31*	*40*	*30*

Table 11.1. Measures reflecting global eye movement performance in CH and the control groups. Mean values are shown, with standard deviations (across trials for CH and across subjects in the control groups).

Fixations directed at the target

In Chapter 3, target-directed eye movements were analysed to see how early and how often targets of different saliency were fixated. Although CH clearly had difficulty performing the task, did she look at the targets? The following measures looked only at those trials where there was a target. To avoid further data loss due to CH's poor task performance, incorrect trials were included.

First, the proportion of trials where the target object was fixated at least once was computed (Figure 11.3). While control subjects fixated targets on the majority of trials (as would be expected with their efficiency in the search task), CH rarely fixated these objects, on only 25% of the trials. This was very different from control group performance (YCs, M=77%, z=2.21, p<.05; AMCs, M=89%, z=5.15, p<.001). A mixed ANOVA looked at the fixation of targets of different saliency. The difference between the control groups was not reliable, $F(1,23)$=2.33, MSE=0.08, p=.14, and the within-subjects effect of saliency was marginally significant, $F(1,23)$=4.55, MSE=0.008, p=.04. There was no interaction, $F(1,23) < 1$. CH was twice as likely to fixate the target if it was medium rather than low saliency, but this was not reliable (Chi-squared test, χ^2=0.89, p=.35).

Given that fixation of the target was rare in the agnosic patient, measures of how early it was fixated are based on only a small number of trials. When fixated, the target was on average reached on the 10th fixation, and this was considerably later than in control subjects for medium saliency targets (mean ordinal fixation on target; CH=12.75; YCs =4.17; z=9.07; AMCs =4.88; z=5.50, both ps<.0001). When searching for low saliency targets CH fixated them later than YCs (7.5 vs. 4.8, z=1.76, p<.05) but within the range of the AMCs (5.98, z=0.79, p=.21). AMCs fixated targets slightly later than YCs, $F(1,23)$=3.17, MSE=3.24, p=.09. Medium saliency targets were fixated reliably earlier than low saliency targets, $F(1,23)$=10.1, MSE=0.94, p<.005, and there was no interaction, $F(1,23) < 1$.

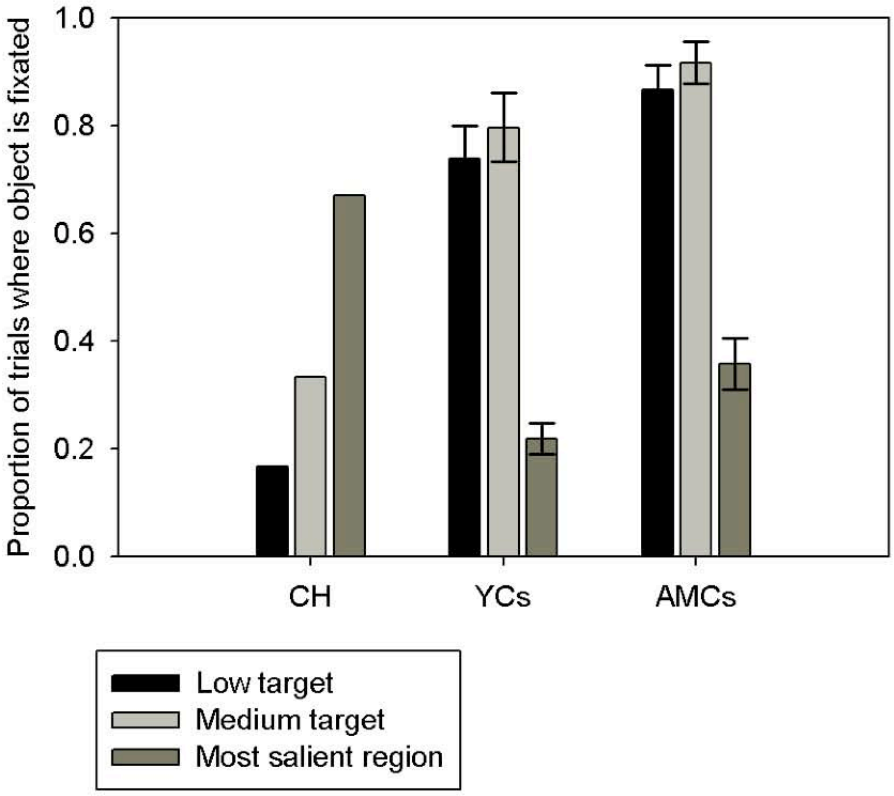


Figure 11.3. The proportion of trials where the target was fixated, as a function of its saliency. The probability of fixating the most salient region in the scene is included for comparison. Error bars for the control groups indicate plus/minus one standard error of the mean.

Fixations directed at the most salient region

Normal participants are efficient when searching scenes and so are rarely distracted by regions based purely on their bottom-up saliency (Chen & Zelinsky, 2006; Chapter 3). The control participants fixated the most salient region in the scene on a minority of trials, though this was more common in AMCs (YCs, M=22%; AMCs, M=36%; $t(23) = 2.61$, $p<.05$; Figure 11.3). It is striking that CH fixated the most salient region much more often (on 67% of all trials; YCs, $z=4.08$, $p<.0001$; AMCs, $z=2.02$, $p<.05$). This comparison is problematic, however: since CH made many more fixations per trial than the control group, there is a higher likelihood that she would fixate the most salient region by chance, even if attention was not being guided by the saliency of scene elements. Of course by this logic then she should also have fixated the target more often too. Nevertheless, to correct for this confounding element an additional analysis was performed.

The frequency of fixations that landed on the two regions of interest—the target and the most salient object—was calculated. Each region was the same size and occupied only .025 of the total image area, so if fixations were distributed uniformly regardless of task or saliency this is the proportion of fixations one would expect on these regions.

The data show an interaction between subject and region of interest (Figure 11.4). The control groups fixated the target with a frequency around 10 times higher than chance, whereas CH rarely fixated the low- and medium-salient targets. In the case of CH, the proportion of fixations on the medium target was also greater than the mean chance expectancy. However, the low saliency target was only fixated very rarely. In all cases CH differed reliably from controls (all zs>2, all ps<.05) and she made more fixations on the target when it was medium saliency than when it was low (χ^2 =14.6, p<.0001). This trend was also reliable in the control groups, F(1,23)=10.54, MSE=0.004, p<.005, though there was no difference between the groups and no interaction, both Fs(1,23) < 1. Thus, as with other measures above, CH fixated the target less frequently than the controls, but did so more often when it was of higher saliency.

The proportion of fixations on the most salient region showed a different pattern. CH fixated this region more often than the control groups (CH=0.11; YCs, M=0.061, z=1.77; AMCs, M=0.066, z=2.14, both ps<.05). The control groups did not differ reliably (t(23)<1).

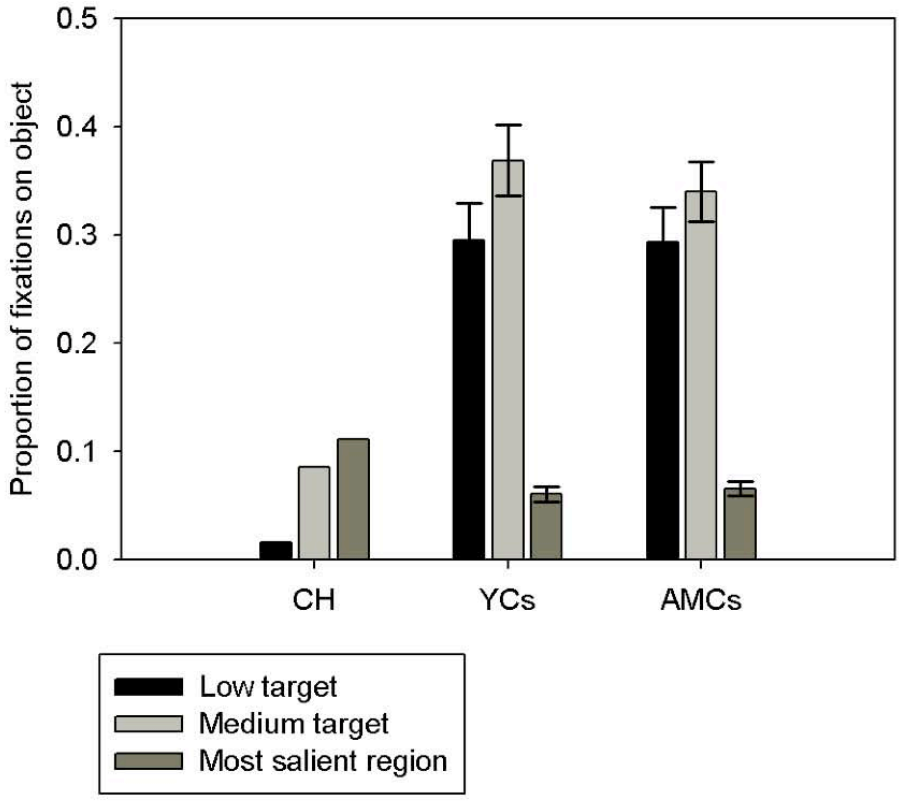


Figure 11.4. The proportion of all fixations landing in each of the regions of interest. Error bars for the control groups indicate plus/minus one standard error of the mean.

Saliency at fixation

CH fixated the most salient region in the scene more often than control groups. However, in order to look more fully at all fixations and regions of different saliency I examined the saliency value at fixated locations. The method for this is described in Chapter 5, and it involved extracting the saliency value at the point in the map corresponding to each fixation location. Assuming a uniform distribution of fixations throughout the scene, a guidance system that selected locations regardless of saliency would lead to a mean saliency at fixation equivalent to the mean of the whole map. As in Peters et al. (2005), I therefore computed the normalised saliency at fixation by subtracting the mean value from that map, and dividing by the standard deviation. An unbiased guidance system would give a mean chance-adjusted saliency not significantly different to zero, while guidance by saliency would be indicated by a reliably positive value. Transforming the

values in this way also gave the advantage of allowing comparisons with different types of stimuli, which will be useful in the next experiment. As discussed in section 5.2, fixations tend to be biased towards the centre of most displays and so comparison with an unbiased map mean will tend to overestimate the relationship between saliency and fixation (Tatler et al., 2005). Here I will focus on differences between patient and control groups rather than getting an absolute estimate of the effect of saliency. As such I will ignore the question of whether the saliency at fixation is meaningfully greater than chance and instead look for a difference between the groups.

The normalised saliency was computed for all fixations with the exception of the first (which was necessarily in the centre) and any fixations that lay on the target region. Target fixations were excluded on the basis that they were specifically primed by the task and that the saliency at their locations was constrained by the experiment. The resulting values were averaged across all stimuli from the experiment, and the means are presented in Figure 11.5. The mean saliency at fixation was larger for CH than for YCs and AMCs, although this comparison was reliable only for the latter group (YCs, z=0.86, p=.19; AMCs, z=1.80, p<.05). The two control groups did not differ ($t(23) < 1$).

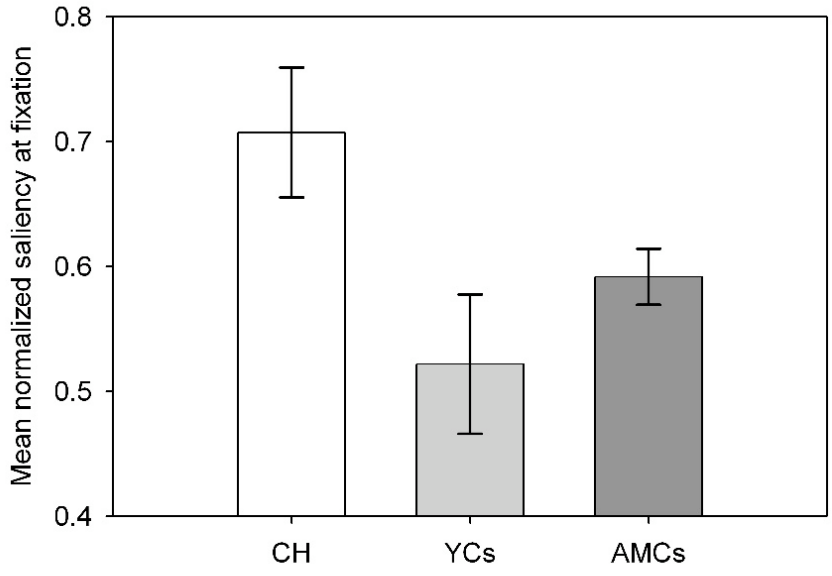


Figure 11.5. The mean, normalised saliency at fixation. Values are averaged across all fixations and trials. Error bars show the standard error of the mean across subjects (in controls) and across trials (in CH).

11.2.3 Discussion

Consistent with neuropsychological testing, CH was severely impaired at detecting objects in natural scenes. She was less accurate and much slower than normal control participants who found the task very easy. Although she scored above chance (in terms of hits) her rate of false alarms, and the fact that she rarely fixated the target and often confused other objects with the fruit targets, is consistent with my overall impression, which is that she was often guessing. Assessment of CH's semantic knowledge suggested that she could produce members of the target category if asked; she knew what fruit were. One can therefore attribute her poor performance to a difficulty in identifying the target visually. It has been suggested that agnosics are particularly poor at recognising objects from a category with high structural similarity between exemplars (Riddoch and Humphreys, 1987). The target fruit were fairly similar to each other, containing round, smooth contours and constant colouring, so this task may have been particularly difficult for the patient (although identifying the exemplar was not technically required by the task). CH's poor search performance is consistent with previous reports of visual agnosia affecting object recognition and conjunction search. The stimuli were fairly realistic which meant that the experiment could investigate the combination of knowledge-driven and image-driven factors on performance and eye movements. Given that a real-world problem experienced by visual agnosics is locating objects around the house this task is particularly apt.

The raw bottom-up saliency of objects does not make much difference in search (Chen & Zelinsky, 2006; Henderson et al., 2007). This is shown in the current study by the fact that control participants rarely looked at the most salient region in the scene (see also Chapter 3). In the YCs in particular there were few effects of target saliency, although AMCs were quicker to respond to medium than to low saliency targets. It is interesting that agnosic performance was also better for targets of higher saliency in most of the measures taken, although this was not always reliable; CH was quicker and more accurate, and was more likely to fixate the target object, when it was medium rather than low saliency. The implication is that with reduced top-down control due to poor recognition of parts of the scene saliency has more of an effect on eye guidance.

The patient rarely fixated the target, so it is perhaps more appropriate to look at the most salient region in the display. Did this capture attention more often in the visual agnosic than in control subjects? There was clear evidence that this was the case; normal controls were able to concentrate most of their fixations on the target, whilst CH was much more likely to fixate the most salient region. She did so on more trials and more fixations than control observers. CH looked at the most salient region more than would be expected by chance or normal performance. Both patient and controls had the same category search instructions, so it is probably not the case that these instructions override saliency automatically. Instead an impoverished representation of likely target features, important in models of realistic search (Navalpakkam & Itti, 2005; Rao et al., 2002), might lead to less weighting of regions likely to contain the target and so more reliance on raw bottom-up saliency. An additional analysis looked at the saliency across all fixations, and all observers showed higher than average values. This suggests that the saliency model was a better predictor of fixations than a uniform strategy. Importantly, there was evidence that values were higher in CH than in controls; her fixations were a closer match to the saliency map.

The use of two control groups allows a high level of confidence that the pattern of results reported is not due to differences in age. In most cases CH was an outlier in comparison to both younger and age-matched controls. There were some differences between YCs and AMCs. Older controls took longer to respond and made more fixations, which were slightly longer in duration on average. It has long been known that ageing can lead to longer response times in visual search (e.g. Rabbit, 1965) and the present research extends this to a more naturalistic task. Other researchers have reported age-related changes in the number of fixations (Scialfa et al., 1994), saccadic accuracy and saccadic reaction time (Munoz, Broughton, Goldring, & Armstrong, 1998) that could explain some of the difference in reaction time. Longer fixation durations in older participants have also been found which could in part be due to differences in the time to program and initiate the next saccade (Scialfa & Joffe, 1997).

Given that this study was concerned with the balance between bottom-up and top-down factors in overt attention it is also interesting to ask whether there are

differences in this balance across age groups. Several authors have suggested that older observers are more affected by distractors or visual clutter and they may even make more "centre-of-gravity" fixations in response to conflicting stimuli (McPhee, Scialfa, Dennis, Ho, & Caird, 2004; Scialfa, Hamaluk, Skaloud, & Pratt, 1999), although this is not always found (Kramer, Hahn, Irwin, & Theeuwes, 1999). AMCs took longer to find targets, but was this due to increased distraction by salient regions? The older observers fixated the most salient region in more trials, but they were no different from YCs in terms of the proportion of all fixations on the target. Thus this difference may be due only to the fact that the AMCs made more fixations overall. There was no difference in terms of the saliency at fixation. Therefore, the differences between old and young controls cannot be confidently attributed to higher reliance on bottom-up cues; CH, on the other hand, showed effects quantitatively and qualitatively different from those of aging, relying far more readily on bottom-up saliency.

I will discuss other aspects of CH's eye movements at the end of the chapter. Before this, the next section describes a second experiment performed with CH looking at recognition for images.

11.3 Experiment 11(b): encoding and recognition of scenes and fractals

CH's performance in a realistic search task confirmed that she could not find objects based on a categorical label. Given the evidence that normal observers perceive the gist of a scene very quickly, how well can CH encode and recognise natural scenes? It has been proposed that gist is extracted from low-spatial frequency information, and that its recognition might occur somewhat separately from the local spatial features which make up objects. Thus it is interesting to ask whether scene processing is impaired in this patient. Visual agnosia can occur in the absence of any memory impairment, and CH showed normal performance on a digit-span test and on other tests of memory. She might, therefore, be able to recognise previously inspected images.

There was some evidence in the previous experiment that saliency was a better predictor of CH's fixations than of controls'. In Experiment 11(b), the procedure from Experiment 2 was replicated with CH and some age-matched control subjects. This

provides a test of CH's old-new recognition performance for natural scenes. Her eye movement behaviour is assessed (at both encoding and recognition) to see whether the patterns observed in search (such as longer fixations and shorter saccades) hold for a different task. In order to explore the impact of image meaning on eye movements in this task, the same procedure was followed with computer generated patterns (fractals) as stimuli. Several authors have used these stimuli because they have similar spatial frequency content to realistic images, and some have argued that saliency is a better predictor of eye movements in fractals than in realistic scenes, presumably because in scenes people are more likely to use their top-down knowledge about the layout of the environment to distribute their attention (Peters et al., 2005; Parkhurst et al., 2002).

With this research in mind several predictions are possible. First, if scene semantics are beneficial for encoding and recognising images, normal observers should be better at recognising scenes than fractals. CH, on the other hand, might be impaired (relative to controls) on scenes, but less so on fractals as accessing semantic details based on visual features is less important in the latter case. In terms of saliency, Parkhurst et al. (2002) would predict a higher correlation between saliency and fixation position in fractals than in scenes. If CH is more susceptible to bottom-up guidance than controls then this correlation should be similar for her in both classes of stimuli, and greater than those of controls.

11.3.1 Method

Participants

As previously, CH is compared to both younger (YCs) and age-matched controls (AMCs). The YC data for natural scenes comes from 21 student volunteers and was reported previously in Experiment 2. A second, separate group of 11 YCs was recruited from the student participant pool at the University of British Columbia and these participants viewed fractals. Two groups of AMCs also took part, viewing scenes (N=5) or fractals (N=4). The AMCs had previously taken part in Experiment 11(a).

Stimuli, design and apparatus

The scene stimuli were exactly the same as those in Experiment 2; there were 45 images seen at encoding and these were presented again with an additional 45 images at test. The same number of computer-generated fractals was taken from the Spanky fractal database, available at www.spanky.net. These were from the same collection as used by Parkhurst et al. (2000) and Peters et al. (2005). Fractals were colour patterns presented at the same size and resolution as the scenes. Figure 11.6 shows some examples of these stimuli.

Figure 11.6. Some examples of the fractal images used in the experiment.

There were two stimuli conditions, scenes and fractals. Patient CH performed the scene condition first, followed by the fractal condition, although she was rested and motivated and was given a break between the two conditions. In the control groups, the factor of stimulus type was manipulated between groups. The eye movement data was further split into three task conditions: encoding, old stimuli at test and new stimuli at test.

As in the previous experiment eye movements were recorded using the EyeLink 1000 (for CH) or the EyeLink II (for all control groups).

Procedure

At the start of the experiment the eye tracker was calibrated to a high level of accuracy and this was repeated where necessary. The experimental procedure was similar to Experiment 2, and was the same for both scenes and fractals. During the encoding phase, 45 images were presented in a random order and the participant was instructed to memorise them. Each picture was displayed for 3000 ms and preceded by a central drift correct marker. Following encoding, 90 images (half from the encoding phase, the remainder novel pictures) were randomly presented in a manner designed to be equivalent to presentation at encoding. Each image was presented for 3000 ms (following a drift correct marker) and this was followed by a screen prompting the participant for a response. At this prompt participants were asked to verbally identify the stimulus as old or new, a response the experimenter logged using the keyboard. This procedure meant that any eye movement artefacts caused by the participant moving to speak were eliminated. It also means that the reaction times are not strictly comparable to the YC scene group (who responded during the 3 seconds; see Experiment 2) so these times will not be analysed. A short practice session consisting of 18 pictures (6 to be encoded followed by 12 at test) was presented before both the scene and the fractals experiment, and these pictures were not presented again in the actual experiment.

11.3.2 Results

The analysis focussed on quantifying CH's behaviour, relative to the control groups, on both scenes and fractals. Hits, false alarms, and overall sensitivity are given as measures of task performance. Following this, global eye movement behaviour is assessed to see if CH's eye movements were different from those made by controls. In terms of saliency, I repeated the salient region analysis from Chapter 5 to see whether CH's fixations were more likely to land on points of high saliency, and I computed the normalised saliency at fixation.

Recognition performance

CH reported that she found the exposure duration very brief but that she did recognise some scenes when she saw them at test. She occasionally made general verbal reports of the contents of the pictures (e.g. "this is outdoors", "this is a house", "a building….looks like a school") although these were not always correct. The proportion of hits and false alarms and a *d* prime measure of sensitivity for all groups is shown in Table 11.2.

		Scenes			Fractals		
		CH	**YCs**	**AMCs**	**CH**	**YCs**	**AMCs**
Hits	***M***	58%	82%	84%	58%	82%	61%
	SD		*9%*	*11%*		*12%*	*8%*
False alarms	***M***	58%	10%	16%	44%	43%	13%
	SD		*6%*	*3%*		*20%*	*10%*
***d* prime**	***M***	0.00	2.33	2.03	0.34	1.20	1.52
	SD		*0.56*	*0.49*		*0.67*	*0.27*

Table 11.2. Recognition data for CH and the control groups. Control data shows means, with standard deviations in parentheses.

Looking first at recognition for scenes, the control groups were relatively good at recognising items, as shown by a large proportion of hits, few false alarms and a relatively high *d*'. CH made fewer hits than controls (YCs, z=2.70; AMCs, z=2.43; both p<.01), more false alarms (YCs, z=7.67; AMCs, z=13.86; both p<.00001) and showed zero sensitivity (significantly different from YCs, z=4.18; AMCs, z=4.16, both p<.0001). In fractals, CH made fewer hits than YCs (z=1.96; p<.05) but was within the normal range for this group on false alarm rate (z=.048) and *d*' (z=1.29). When compared to the AMCs, CH made around the same number of hits (z=0.40) but many more false alarms (z=3.17, p<.001) and with a lower sensitivity (z=4.41, p<.0001). Thus CH was particularly impaired on the scene recognition task, but somewhat less so on the fractals. There were only slight differences between the control groups, and given the small number of AMCs these were not analysed further. Instead, the control groups were combined to explore the effect of stimulus type on overall sensitivity. A one-way

ANOVA showed a significant effect, $F(1,39)=28.6$, $MSE=0.32$, $p<.0001$; control subjects' performance was better for scenes than fractals. CH showed no such benefit and in fact was slightly better for fractals.

Global eye movement measures

The number of fixations, the average fixation duration and the mean saccadic amplitude was computed for each trial, and these are shown in Table 11.3. In scenes, CH made more fixations than YCs during old and new trials at test (both zs>4.5, $p<.0001$), but she was within normal limits at encoding ($z=0.1$,$p=.45$), and in all cases did not differ reliably from the AMCs (all $z<1.6$, $p>.05$). In fractals, the number of fixations made by CH was more similar to the YCs and none of the z-scores were reliable (all $z<.07$, $p>.20$). During the test phase, CH made reliably fewer fixations than the AMCs ($z=1.67$ and 1.90 for old and new items respectively, both $p<.05$). There was no difference during encoding ($z=0.16$, $p=.44$).

As there was a fixed viewing time the mean fixation duration was inversely related to the number of fixations. In both scenes and fractals, CH tended to make shorter fixations, on average, than controls. However none of the comparisons were reliable (all $z<1.1$, $p>.10$).

The most striking observation, which is consistent with that seen in the previous experiment, is that CH made far shorter saccades than the control subjects. In fractals, this was reliable in all types of trial and when compared to both YCs and AMCs (all zs at least 1.8, all ps<.05). Whilst viewing scenes, her saccades were shorter than YCs in all phases ($z=5.0$, 2.1 and 2.8 for encoding, old items and new items respectively, all $p<.05$), and they were also shorter than AMCs in encoding trials ($z=2.9$, $p<.005$), and in new trials at test ($z=2.0$, $p<.05$). In old trials this comparison was not reliable ($z=1.1$, $p=.13$).

			Scenes			Fractals		
			CH	**YCs**	**AMCs**	**CH**	**YCs**	**AMCs**
Number of fixations per trial	**Encoding**	***M***	10.40	10.60	10.70	10.30	9.40	10.00
		SD	*1.47*	*1.40*	*0.65*	*1.50*	*1.41*	*2.05*
	Old	***M***	16.20	10.80	14.60	10.10	10.00	12.80
		SD	*5.68*	*1.16*	*2.37*	*1.38*	*1.16*	*1.62*
	New	***M***	17.90	10.50	14.60	9.90	10.00	12.80
		SD	*12.77*	*1.26*	*2.10*	*1.29*	*1.38*	*1.53*
Fixation duration (ms)	**Encoding**	***M***	248	278	247	231	314	303
		SD	*52*	*57*	*31*	*51*	*77*	*115*
	Old	***M***	234	264	229	272	283	245
		SD	*48*	*38*	*29*	*51*	*47*	*38*
	New	***M***	235	279	229	270	292	244
		SD	*35*	*49*	*28*	*48*	*66*	*33*
Saccadic amplitude (˚)	**Encoding**	***M***	3.5	6.2	6.5	2.6	7.8	5.9
		SD	*0.97*	*0.54*	*1.03*	*1.05*	*1.71*	*1.81*
	Old	***M***	4.7	6.3	6	3.8	8.9	6.4
		SD	*1.35*	*0.79*	*1.16*	*1.28*	*1.84*	*0.84*
	New	***M***	4.4	6.7	6.3	3.7	8.9	6.3
		SD	*1.22*	*0.8*	*0.96*	*1.25*	*1.71*	*0.97*

Table 11.3. Global oculomotor statistics from CH and the control subjects in scenes and fractals. Cells show the mean and standard deviation, calculated across trials in CH and across participants in the control groups.

How did the type of image affect viewing behaviour? As previously there were too few AMCs to perform parametric statistics, so the control groups were collapsed across age and entered into a 2-way ANOVA with the within-participants factor of trial type (encoding, old items and new items) and the between-groups factor of image type.

The number of fixations was not affected by image type, $F(1,39)=2.13$, $MSE=8.19$, $p=.15$, but there was a reliable effect of trial type, $F(2,78)=12.99$, $MSE=0.991$, $p=.001$, with fewer fixations at encoding than either type of trial at test (both $t(40)>3$, $p<.01$). There was no interaction, $F(2,78)<1$. As one would expect given the fixed viewing time, mean fixation duration was also affected by trial type, $F(2,78)=8.61$, $MSE=815.21$, $p<.005$, and was shorter in old images (both $t(40)>3$, $p<.005$). Image type did not affect the mean fixation duration reliably, $F(1,39)=1.74$, $MSE=7451$, $p=.20$, and neither did it interact with trial type, $F(2,78)=2.70$, $MSE=815.2$, $p=.07$.

The variation in saccadic amplitude was more interesting: longer saccades were made in fractals than in scenes, $F(1,39)=14.31$, $MSE=4.74$, $p=.001$. Saccade length also varied between the different trial types, $F(2,78)=15.83$, $MSE=0.221$, $p<.001$, and these effects were qualified with an interaction, $F(2,78)=9.54$, $MSE=0.221$, $p<.001$. Saccades were larger in fractals across the whole experiment, but this was most pronounced in the test phase (simple main effects of image type: at encoding, $F(1,117)=6.16$, $MSE=1.73$, $p<.05$; at old, $F(1,117)=21.97$, $MSE=1.73$, $p<.001$; at new, $F(1,117)=13.58$, $MSE=1.73$, $p<.001$.

Fixations on salient regions

The analysis in this section was directed at comparing the match between fixation patterns and the saliency map, in different types of image and in CH and controls. I presented an in-depth analysis of this for YCs with scenes in Chapter 5. Here, I repeated an analysis of the proportion of fixations that landed on one of the first 5 most salient regions, according to the model. There was not much of a difference between the different task phases in this respect in Experiment 2, so here all trials were combined in this analysis. In scenes, an average of 18.3% (SD=2.3%) of all the fixations made by YCs were located in one of the top five regions. This value was slightly lower for the AMCs (M=16.9%, SD=2.0%) and for CH (M=15.2%), although the patient was within the normal range ($z=1.35$, $p=.09$ and $z=0.84$. $p=.20$ when compared to YCs and AMCs respectively). In fractals, control participants spent a similar proportion of fixations in the five most salient areas (YCs, M=16.7%, SD=2.3%; AMCs, M=16.7%, SD=1.4%). CH actually made fewer fixations in salient areas (M=12.5%; vs. YCs, $z=1.83$, $p<.05$; vs.

AMCs, z=3.01, p<.005). An independent-samples t test compared the proportion of control participants' fixations on salient regions in scenes and in fractals. There was a slightly higher proportion of fixations in salient areas in scenes than in fractals, but this difference was only marginally reliable, $t(39)$=1.94, p=.06.

Saliency at fixation

As in the previous experiment, I took the normalised saliency at fixation and averaged across all fixations following the first saccade, in both the scenes and the fractals. The means for CH and the control groups are shown in Figure 11.7.

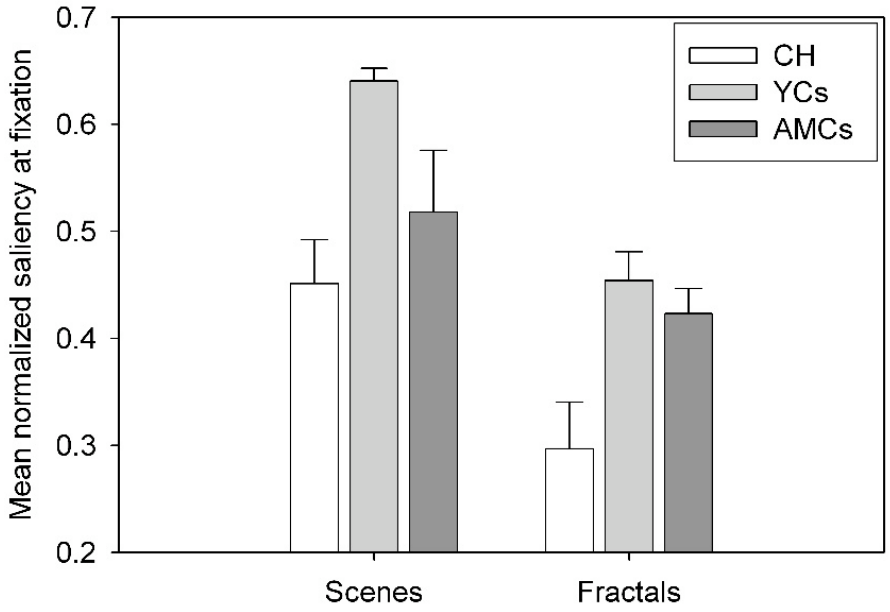


Figure 11.7. The mean normalised saliency at fixation, for different types of stimuli in the recognition task.

In contrast to the results from the search task, CH actually showed lower saliency at fixation than control groups. She was outside the normal range when compared to the YCs (scenes, z=3.43,p<.0005; fractals, z=1.76, p<.05), but was reliably different from the AMCs only in fractals (z=2.65, p<.005) and not in scenes (z=0.51, p=.30). Combining the control groups, fixations in scenes showed a reliably higher normalised saliency, $t(38)$=6.25, p<.001. This is contrary to what was observed by Parkhurst et al. (2002).

Eye movement biases in scenes and fractals

In contrast to the previous experiment there was no evidence that CH looked at salient regions more often than controls in this task. In Chapters 5 and 10 I argued that image-general eye movement biases should be taken into account when comparing saliency and fixation. If these biases differ in agnosia, and in different pictures, then this will have an impact on any relationship between saliency and fixation.

Inspecting the overall distribution of eye movements suggested that CH's fixations tended to be more centralised than control subjects'. To quantify this, the mean Euclidian distance between all fixation locations (not including the first) and the centre of the screen was computed. Consistent with my observations, CH's fixations were consistently closer to the centre of the display. In scenes, the mean distance from the centre was 5.39˚ for CH, compared with 7.72˚ for YCs (SD=0.56, z=4.12, p<.0001) and 7.35˚ for AMCs (SD=0.72, z=2.72, p<.005). Agnosic behaviour was also outside the normal range on this measure in fractals. In these trials, CH's fixations were 3.93˚ from the centre, on average, and this was reliably closer than both YCs (M=6.6˚, SD=0.89; z=2.98, p<.005) and AMCs (M=5.8˚, SD=0.93; z=2.01, p<.05). Thus, in both types of stimuli, CH tended to fixate closer to the centre than controls, and indeed this is perhaps what would be expected given that viewing started in the centre and that CH made shorter saccades. These means also illustrate that, in all groups, fixations were closer to the centre in fractals than in scenes. A t test of this difference between the control groups was reliable (t(39)=5.30, p<.001).

The previous chapter reported a pronounced horizontal bias for saccades in natural scenes. The use of the fractals here gave an opportunity to examine saccade direction bias in a different type of stimulus that had no obvious structure. This analysis was carried out on all the YCs saccades in fractal images, excluding those that were less than 1˚ in amplitude. As previously, saccades were grouped according to their direction into four symmetrical bins equivalent to horizontal saccades, vertical saccades and those at oblique vectors. The results indicated that there was indeed a bias for saccades in the horizontal direction. On average, 34.8% of a participant's saccades were within 22.5˚ of the horizontal axis (SD=11.2%), compared with 19.1% (SD=2.8%), 23.0% (SD=11.8%)

and 23.1% (*SD*=4.3%) in the 45˚, 90˚ and 135˚ axes respectively. Repeated-measures ANOVA showed a reliable effect of axis on saccade frequency, $F(3,30)=6.09$, $MSE=10656$, $p<.005$. There were more saccades in the horizontal axis than in the mean of the other axes ($t(10)=3.09$, $p=.01$). Given the relatively smaller number of participants and saccades available in CH and the AMCs, these data were not analysed in detail. However, the horizontal bias was similar in all participants, including CH, who made 31.9% of saccades in the horizontal axis compared with 25.3%, on average, in each of the other three bins.

11.3.3 Discussion

This experiment built on the observations from a search task and asked whether CH could recognise scenes and fractals, and whether her allocation of fixations was anomalous when encoding or recognising these images. CH found this task very difficult, and thus showed recognition performance that was close to chance, and much poorer than control subjects. Previous testing of CH indicated that she had normal short- and long-term memory. A more extensive evaluation of CH's visual memory would be necessary to make firmer conclusions, but it seems that her impediment in this case was not a failure of storage, but rather one of encoding and retrieval. Considering her difficulty in recognising the items she looked at, she would be unable to encode them semantically. At test, even if she had stored some information about the images she had seen, her recognition of these features in the current array would have been slow and inefficient. In her case report it is noted that this patient performed poorly at recognising famous faces, which would include those she had encountered earlier in her life when one assumes her feature recognition was normal. Therefore one supposes that her deficit is also one of retrieval.

There are two interesting points about CH's performance in this task that would be interesting to pursue in further research. First, to what degree could CH's performance be improved by encouraging eye movements more similar to those of controls? One of the reasons that CH found the task so difficult was that there was a relatively brief time limit in which to explore the pictures. Normal participants (including those of the same age) performed well under this pressure, but it is likely that the patient's performance

would improve if she were given more time. The time available for encoding and recognising could be manipulated, and this might give some information about the cause of CH's difficulties. Alternatively, she might be encouraged to make larger saccades and this might enhance her performance.

Second, how does the type of image affect strategies at encoding and recognition? If CH's perception of simple visual features is relatively normal, she might be expected to do better with stimuli such as fractals that do not require semantic identification. The control subjects performed somewhat better at recognising previously seen pictures when they were meaningful scenes than when they were fractals. This might be because they could encode the scenes more efficiently by incorporating their previous knowledge. On the other hand, it was hard to control for the similarity of old and new items in this test and it may have been that the recognition test with fractals was actually more difficult because they were more similar. CH performed very poorly on both types of stimuli so there was no evidence under these conditions that she did better with more abstract images.

11.4 General discussion

In both the experiments in this chapter, there were large differences in the eye movements made by an agnosic patient and controls, and these were mostly the same with both student and age-matched controls. In search, CH made more fixations, dwelt in one location for longer (as shown by a longer average fixation duration), and made shorter saccades. Impairment in object perception had dramatic effects, not just on performance but also on the eye movements made whilst searching. This supports a large role for top-down guidance in the eye movements of normal populations during search and scene perception, and this guidance allowed controls to move quickly to the target. The higher frequency of fixations and the shorter saccades made by the patient may reflect the difficulty in acquiring information when normal object recognition processes do not guide perception. The longer fixation durations indicate that she took longer to process objects and other details (presumably because she was trying to identify them and work out whether they were targets) while control participants could very quickly reject non-targets and move on. Assessment of CH's low-level vision was normal; she was able to detect

targets quickly in the peripheral field and within the range of normal observers. It therefore seems likely that the differences found in global eye movements are due, not to bottom-up processing difficulties but a top-down recognition deficit.

In the recognition task, where all participants had a fixed time limit, CH's eye movements were more similar to those made by normal observers. When recognising scenes, she made more fixations than younger, but not age-matched participants. Her fixations were slightly lower than normal, but not reliably so, and conclusions about these measures are complicated by the requirement to respond during the trial. This finding does show, however, that the differences seen in search are not because CH cannot make fixations at the normal rate under any circumstances, but rather that they depend on the task at hand.

Shorter saccades in the agnosic patient were observed in both the experiments, and in both scenes and fractals. The robustness of this finding suggests that it would be fruitful to look at this further. It was assumed that the patient could make long saccades if necessary and neuropsychological testing with simple stimuli confirmed this. A higher level explanation for CH's longer saccades might be that, due to her difficulty in recognising objects, she refixates objects more often than normal controls, leading to many smaller, within-object shifts.

Were the eye movements of the agnosic patient preferentially directed towards areas of high, bottom-up saliency? In the search experiment there was some evidence that, rather than fixating randomly across the display, CH was more likely to fixate salient regions than were controls. However, in the recognition experiment, CH actually spent slightly less time fixating on the most salient regions. Although this is contrary to my predictions, the discrepancy may be resolved by considering the different tasks. In Chapter 3, and using the same stimuli for both tasks, saliency was more important in an encoding task than in a search task. Applying this finding to the current results, it appears that in a search task the behaviour of controls was to neglect salient regions and look to the target, and without top-down recognition CH's eye movements were qualitatively different. In an encoding task, Experiment 1 suggested that participants looked at more salient objects more often. If there is no particular reason to look at other objects then

guidance toward salient regions might be as good a strategy as any in this task. If this is the case, then top-down guidance may coincide with saliency-driven eye movements in the memory task, and so perhaps it is not surprising that the patient and controls do not differ here in terms of their tendency to fixate salient regions.

One should be cautious about concluding that the saliency model can predict fixations better than chance in these tasks. The aim of this chapter was to consider differences between the patient and controls, and so I have not focussed on making the detailed comparisons with a random model that were discussed in Chapter 5. A potential complicating factor is that the distribution of fixations across all images was different for CH and normal participants. The centralisation bias seen in other studies was more pronounced in CH.

11.5 Conclusions and links forward

I have made novel observations regarding scanning in visual agnosia and these suggest not just an increase in task difficulty but a shift in the importance of bottom-up guidance which manifests itself in eye movement behaviour. It appears that our patient employed a degree of top-down control in her desire to complete the task successfully, however her agnosia gave rise to the selection of areas for inspection based more upon raw saliency than prior knowledge of object form which can be utilised by controls. Consequently her visual search was laborious and dramatically extended in time, dwelling, as she did, on objects in the scene inordinately longer than a healthy population in an attempt to mine meaning from them. In an encoding task, there was no greater correlation with saliency in agnosia, and this appears to be due to the task demands.

12 General conclusions: towards a model of eye guidance in natural scenes

The experiments described in this thesis aimed to clarify the role bottom-up saliency plays in where people look in natural scenes. In particular, they allow the model proposed by Itti and Koch (2000) to be evaluated. This chapter begins by reviewing the key findings from the previous chapters. What can the model predict about eye movements in different situations? I will then draw on other findings from this thesis to suggest how a model of eye movements in natural scenes can be refined.

12.1 Summary of main findings

12.1.1 Where does the model succeed?

There were reliable effects of saliency in almost all of the experiments, and these were seen in the eye movements leading to fixation of a region, the processing of this region once fixated, and the performance of the experimental task.

The model is designed to explain how fixated regions are selected and it does so by evaluating their conspicuity based on biologically plausible features. In tests of this, Experiments 1, 3, 4 and 5 all showed that objects with higher model-predicted saliency (those which were selected earlier by the model) were fixated *earlier* compared to other regions. This was true in a range of situations, both when the region was directly relevant to the task (e.g. when it was a target in Experiment 3) and when it was not. In a complex scene with many potential items competing for attention, salient regions were also *more likely* to be fixated than one would expect by chance. In Experiment 1, salient objects were also refixated more often. In two experiments, there was evidence that eye movements to salient regions came from *further* away, suggesting that saliency affects the extent to which items in the periphery can attract saccades. There remains a paucity of experimental evidence testing the model so these findings are important. An alternative way of evaluating the model is to compare the saliency at fixation with that from a random sample. This comparison in Chapter 5 confirmed what was found by Parkhurst et al., (2002) and Peters et al., (2005): that saliency at fixation is higher than chance, even when a biased estimate was used.

Many models of eye movements treat the decision about where to move the eyes separately from the decision about how long to stay in one location. This follows from the assumption that the processing time during a fixation relates to the information at that spatial location. There is, then, no reason why the saliency model should predict inspection durations. However, several results in this thesis showed that salient regions were in fact fixated for longer and returned to more often, although this was not always the case. I argued in Chapter 7 that this is because saliency is correlated with semantic informativeness. In the experiments investigating search for different scene regions, the stimuli were relatively natural and uncontrolled so the meaning of the regions was free to vary. In those experiments where specific objects were manipulated the same object could be made more or less salient. For example in Experiment 1 the objects of interest were always pieces of fruit (and sometimes the target), so it is not surprising that in these cases there was no difference in the length these items were inspected. It is also interesting that fixation duration was affected by peripheral filtering in Experiment 5. If the process of choosing a new saccade target affects the length of the current fixation then the saliency of the next location might also affect this.

Where people pay attention and move their eyes to affects their ability to perform different tasks, under some conditions. For this reason saliency also had an effect on task performance. For example, people responded to indicate that they had found a salient region quicker and more accurately than a less salient region (Experiments 4 and 5). When the salient object is irrelevant to the task it may distract the observer and lead to a detriment in performance, even when it is not fixated (Experiment 9). In Chapter 11 there was some evidence that the fixations of an agnosic patient in a search task were more closely tied to the saliency map than those of controls. This is support for the approach and suggests it might be extended to investigate neuropsychological deficits in scene viewing.

12.1.2 Where does the model fail?

The previous section shows that the model does have some utility in predicting eye movements and resulting behaviour. However there were many limitations, both in terms of the accuracy of the predictions and the situations where the model could not explain the observed data.

First, although large differences in saliency rank produced reliable differences in behaviour the model was not accurate at predicting exactly when a region would be fixated. The first most salient point according to the model was not normally the first point to be fixated, even though it may have been fixated more often than chance. In Experiments 3,4 and 5 the fifth most salient point was not consistently selected any more than random control regions. Another way of thinking about this is to look at the sequence of fixations predicted by the model. In Chapters 5 and 8 the model sequence was not reliably similar to the order of fixations made by participants. It may be that the model's efficacy breaks down after the most salient region (i.e. the first predicted shift), although the target objects in Experiment 1 were less salient than this and still led to reliable effects. The saliency map approach proposes not just an algorithm to mark out important regions but a dynamic system that will move through the scene. However, there was little evidence that the order of salient regions in the scene was related to fixation patterns. If the eye movement system selects points of gradually decreasing saliency, then saliency at later fixation locations should be lower, but this was not the case in Experiment 2.

Second, effects of saliency were not always found, and in search in particular there were cases when saliency was dominated by the task. A strong version of the saliency map model predicts that people should fixate salient regions preferentially, regardless of the task. Such a hypothesis is untenable. This was clearly illustrated in Experiment 1 where saliency had a reliable effect in a memory-encoding task but not with the exact same stimuli in a search task. Targets were fixated quickly regardless of their saliency. On the other hand, when a time limit was introduced there were differences in the efficiency by which salient targets could be found (Experiment 1(b)). In other search

tasks, there were some effects of saliency, although the semantic meaning of the regions may have been just as important. When more controlled arrays of objects were used, target saliency had some small, though reliable, effects on eye movements, as did the saliency of a distractor. It seems, therefore, that search instructions do not completely override bottom-up saliency as a guidance factor, but they do obscure it and there was a lot of evidence for top-down control.

In Chapters 8 and 9, despite some effects on eye movements, the saliency model was not good at predicting the order in which objects were fixated in a search task, even when the target was not present. Specifically, in Chapter 8 the model- and participant-generated scanpaths were not similar. In Chapter 9, there was no relationship between saliency and probability of inspection in target absent trials. This can be interpreted in terms of the two problems discussed above. The saliency map model may be above chance at predicting the general areas that will receive fixation (and those that will not), but it is not efficient at predicting movements between them. In addition, salient items do not invariably capture the search process. Instead, participants can quickly and easily translate a verbal or pictorial target into a set of features that can bias search. In future, implemented top-down models should be used to quantify target-item similarity, and this will predict the search path better than a purely bottom-up model.

Although the saliency at fixation locations is higher than chance, I showed in Chapter 5 that taking into account general eye movement biases reduces this difference. In other words, the correlation with saliency depends completely on which random model one compares it with. If some of the habitual patterns observed in this thesis, in particular the tendencies to fixate centrally and to make horizontal saccades, are not dependant on the saliency distribution in the image then they might explain just as much variance in where people look as the saliency model can (or even more). Clearly the saliency model is better than a completely uniform model with no assumptions about where in a scene people are likely to fixate. However, in natural images this is too liberal a comparison.

12.1.3 What other factors are important?

As mentioned, the centralised pattern of fixations, and the biases in saccade direction are particularly robust biases that need to be explained by any model. As several recent saliency map implementations have pointed out the variation in visual resolution with eccentricity needs to be taken into account (Vincent et al., 2007). Down-sampling the information at peripheral locations might go some way towards explaining a central bias (by promoting shorter saccades) as well as the finding in Experiment 9 that objects near the edge of the screen were rarely fixated.

However, it also seems likely that a systematic, head- or scene-centred strategy to look toward the centre is used by participants, particularly on the very first saccade on seeing an image. This is most likely a purely top-down strategy that would occur regardless of the saliency content of the image (although it may have been learnt over exposure to images which tend to show a central bias). The pattern of within-object landing position seen in Experiment 9 is a more local form of eye guidance that needs to be explained. It would be useful for future research to see if this pattern is found in complex displays and real scenes. Within a saliency map framework, this could be modelled by computing some kind of average across the surface of an object that would result in guidance to the centre. It is interesting to note that a bottom-up approach is likely to favour saccades to the edges of an object (where there is the most contrast), which is not the pattern that was found in Experiment 9. This is fundamentally an issue with whether attention and saccades are allocated towards regions of space or objects. Perhaps the solution for both the scene-centred and the object-centred biases would be that saliency is combined with a coarse representation of layout which parses a scene into objects.

Across experiments, participants tended to fixate those areas that were semantically relevant, either because they were targets in a search, or because they were areas which would be useful to encode or understand. Unfortunately there are few quantitative models for determining this in a natural scene. Saliency is correlated with semantic informativeness, and this is what makes it plausible as a guidance factor.

However, experiments are necessary which separate these factors. In many cases saliency would be a useful visual short cut for when higher-level meaning is inaccessible (e.g. due to limits in peripheral processing). There is some evidence in this thesis that, when knowledge about scene regions is limited (as in visual agnosia), or when task demands are high (for example when there is a restrictive time limit as in Experiment 1(b)), saliency becomes a better predictor.

12.2 A framework for eye movements in scenes

The findings discussed above identify the different processes that contribute towards eye guidance in natural scenes. Saliency had an experimental effect in several cases, and one advantage of including it in a model of eye movements in scenes is that it appears to contribute in several different task situations. On the other hand I have argued that it is heavily moderated by task and scene knowledge. With the aim of making these processes more explicit, I will now describe a framework for eye movements in natural scenes.

Figure 12.1 illustrates the general model, and I will describe each component in turn. The scheme combines a saliency map with spatial priors and a spatiotopic representation of the top-down relevance of different scene regions. I will then consider how this framework explains the results in both encoding and search tasks.

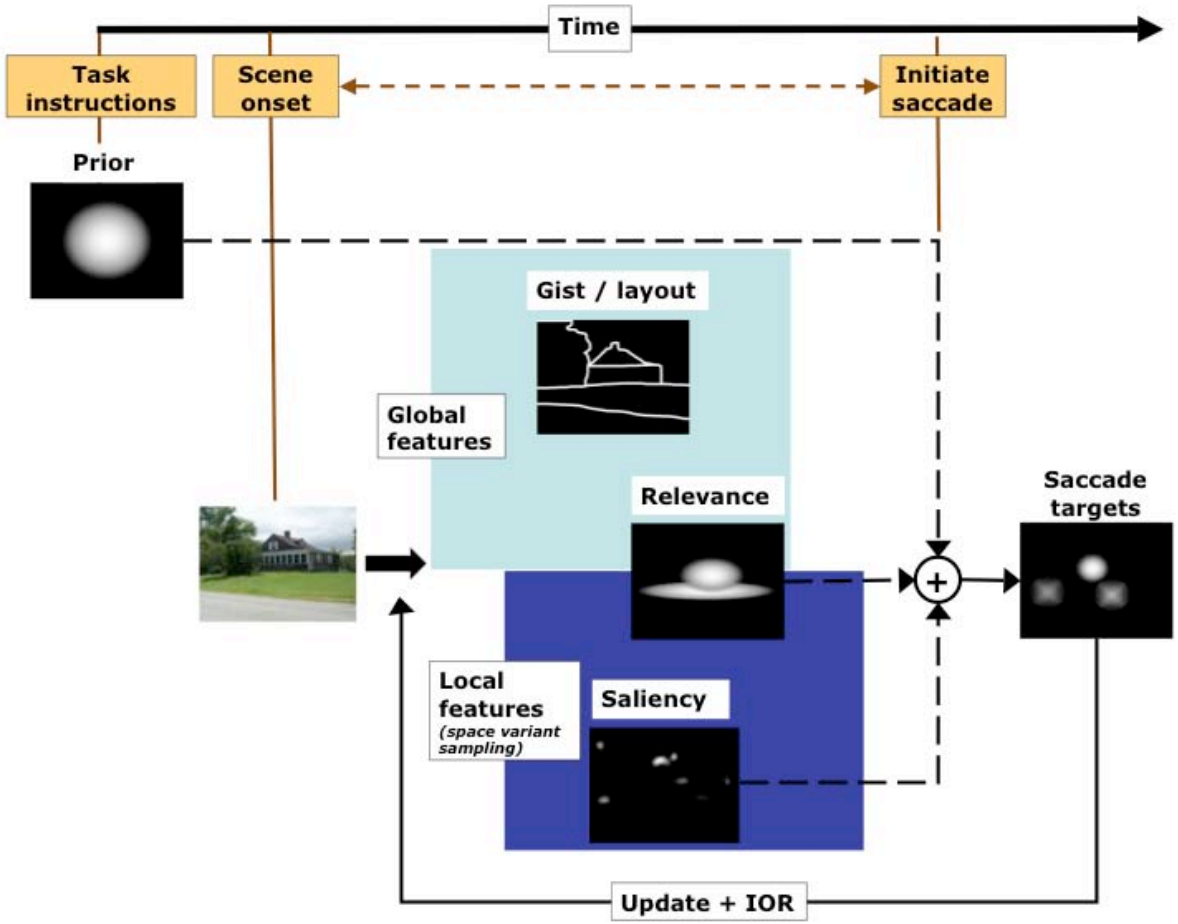


Figure 12.1. A framework for explaining eye movements in natural scenes. Processes are organised over time, beginning at the far left when the scene has yet to appear. The potency of each part of the scene to attract attention is represented in several spatiotopic maps with brighter areas indicating points more likely to be fixated.

On receiving task instructions, a prior is formed representing expectations of where important features are likely to be. When the scene is onset gist and layout are processed holistically, whilst local feature saliency is computed in parallel. A relevance map represents the knowledge-driven importance of parts of the scene, based on shared features with the target and task-relevant locations.

Priors, relevance and saliency are summed into a single map which controls saccade targeting.

Unlike a pure saliency model, the first (leftmost) stage in Figure 12.1 incorporates the expectations of the task in terms of a spatial *prior*. I have argued that a tendency to fixate centrally occurs, at least some of the time, independently of the features present in the scene. One such prior, therefore, might be to automatically orient towards the centre, and this preference would be present before the scene appears. In Experiment 10 people tended to move to the centre on the very first saccade and it is assumed that this was a top-down strategy driven by the participants' knowledge of where the image would appear. In a search task the instructions would include the identity of the target, and this might include an expectation of where it will occur which could also be instantiated in this prior. For example, if the search target were a chimney, participants would be biased toward the top of the screen before the to-be-searched scene had even appeared. Thus, the prior weights certain locations based on top-down expectations provided by task and target before the scene is perceived. If there were no explicit task or expectations all locations could be equally (un)important (although in this case a default central strategy might well apply).

I have divided the features processed when perceiving the scene into two parallel streams dealing with global and local features, with different time courses. Torralba et al. (2006) also distinguish between global and local pathways, although they do not explicitly discuss the speed of the processes except to say that they combine prior to image exploration. In the present model, global features are processed quickly and they provide the initial scene gist and layout. It is known that scene category can be identified at extremely short exposures and research has shown that this context can be apprehended in a holistic fashion, independently of focused attention and without the need to recognise the constituent objects (Biederman et al., 1974; Oliva & Torralba, 2007; Potter, 1976). In the context of eye movement control the gist and layout information, in concert with task expectations, will contribute to the parts of the scene considered most relevant and worthy of fixation. Experiment 10 suggested that the location of the horizon was a particularly important cue for biasing saccade direction. I propose that this orientation information is extracted from the global features at this point.

The local features channel proceeds in parallel to the global pathway and consists of the analysis of local image features, as in the saliency map model. Local features are present at high spatial scales, and thus they are subject to the constraints of the retina, namely the decrease in resolution with eccentricity. For this reason the present scheme follows Vincent et al. (2007) and proposes biologically plausible, space-variant sampling of local features, so that regions at the fovea are more likely to be salient than those in the periphery. The saliency map is computed as in the Itti and Koch (2000) model, with locations becoming salient if they contrast with their surround.

In the framework in Figure 12.1 the target for the next saccade is determined not just by bottom-up saliency but by a *relevance* map which represents the importance of different regions to the task. Relevance is computed top-down, based on the correlation between those features present in the image and those deemed important for the task (e.g. those belonging to the target). This appearance-based target weighting is implemented in Rao et al.'s (2002) model of object search but here I also include contextual information based on the gist. For example, relevant locations in a landscape would be located near the horizon. If there is an explicit target then the gist of the scene will bring with it expectations of where the target will appear.

The different priority maps are summed into a single saccade-targeting map that represents where the next fixation will be located. Fixation location will therefore be determined by prior expectations of task and target, local saliency and the relevance computation. As these maps are summed, salient areas may continue to attract eye movements even if they are not relevant for the task. This detail differs from the contextual guidance model of Torralba et al. (2006) who suggest that areas that are salient but inconsistent with the context are vetoed completely. Once the saccade has been executed, the local feature channel is updated to reflect information acquired in the fixation. This would also include the inhibition of return present in the saliency map model that reduces the saliency of the previous saccade target.

A final noteworthy feature of this framework is that the time taken to initiate a saccade interacts with the evolution of the different maps. While this is only a tentative suggestion, given the different time-courses of the gist, relevance and saliency

computations it can explain some of the results from this thesis. For example, in Experiment 1(a) there was no effect of target saliency in a search task. In the current framework this is explained by top-down relevance dominating eye guidance. However, when there was a time limit (Experiment 1(b)) saliency did have an effect. If it is assumed that computing relevance based on a complex conjunction of target features takes time then the necessity to initiate saccades earlier might mean that the relevance map has not fully evolved and so saliency will become more important. On the other hand, if the computations involved in the different maps are delayed this may have an effect on the time to initiate a saccade, therefore prolonging the fixation duration. This would explain the differences in fixation duration found in Experiment 5 where filtering of peripheral information affected the time taken to move the eyes.

Although not fully implemented, this framework is useful in discussing the findings from this thesis. In the following sections I will discuss how it can be applied to two different experimental tasks: general scene encoding, and object search.

12.2.1 Explaining general scene viewing

Several of the experiments in this thesis can be characterised as requiring a general encoding strategy. In Experiments 1(a), 2 and 3 participants had to view pictures in preparation for a memory test, whilst in Experiment 10 they were subsequently tested with a sentence verification task. In each case, although there was no specific target, eye movements would have been guided both by preconceptions and understanding of the scene and by visual saliency.

In Experiment 2 it was observed that people were much more likely to fixate in the centre of the screen than elsewhere, although this was confounded by the fact that viewing started at this point. In Experiment 10, when fixation at scene onset was at a peripheral location participants almost always moved to the centre of the image. This behaviour is explained by the current framework as a spatial prior or preconception about where useful information is likely to occur. This prior might also prime locations along the picture horizontal, which would explain the predominance of horizontal saccades found in scene viewing. However, in Chapter 10 the distribution of saccade directions changed when the picture was rotated. The pattern is therefore better explained by the

acquisition of gist and top-down guidance towards areas of expected relevance. Assuming that scene orientation is one of the attributes that can be acquired from global features, the relevance map might be updated to bias attention towards areas near the horizon. If the scene were rotated 90˚, saccade targets on the horizontal would be primed.

How does the framework outlined above clarify the effects of saliency in an encoding task? Experiments 2 and 3 showed that people were more likely to fixate salient regions than would be expected by chance. These regions would have been represented by higher activation in the saliency map, and this activation would feed through to the targeting map. They may also have been more informative, and this would increase their potential of being fixated further via the relevance map. In the memory condition in Experiment 1, the regions of interest were similar objects and so were presumably equally relevant to the task. The fact that objects which were ranked as more salient by the Itti and Koch (2000) model were fixated earlier and more often is good evidence that saliency is added to relevance when deciding on where to fixate. Another situation where one might expect the influence of relevance and gist to be reduced, and thus for saliency to dominate, is in the inspection of fractals. However, in Chapter 11 there was no evidence that fixation patterns were better related to saliency in these images.

12.2.2 Explaining natural visual search

The degree to which bottom-up and top-down factors interact in visual search continues to be a topic of interest for researchers (Zelinsky et al., 2005). The framework I propose includes both image-dependant and knowledge-dependant influences on naturalistic search and I will now summarise how it explains the findings in this thesis and elsewhere.

It is clear that people are able to relatively easily bias their search and their eye movements away from salient regions and towards the target. This was seen in Experiments 1, 4, 6 and 7, where people found low saliency targets just as quickly as higher saliency ones. In the model described above this top-down guidance proceeds via an early prior (which represents preconceptions about where an object will appear) and a relevance computation. In the first instance the target template may provide frame-centred clues about where an object will appear. This helps to explain why some regions

were found earlier than others in Chapters 6 and 7. In these experiments, salient regions tended to contain objects and features, and recognising these provided some information about their likely location within the frame of the image, before the specific image even appeared. If the target were a bird or a patch of sky (and therefore most likely to appear in the top of the screen), this prior would be different from if it were a person or a patch of grass. Within search arrays where target location was less constrained, as in Experiments 6, 7 and 9, the prior is assumed to reflect the real probabilities and therefore to be flat, with all object locations equally likely to be fixated.

Looking at the relevance computation in natural search, various models describe the way in which features from the target are compared to those at each point in the scene in order to generate a target probability at that location (Navalpakkam & Itti, 2005; Rao et al., 2002; Wolfe, 1994). In my framework this is combined with contextual information from the gist of the image to prime likely locations. Research has shown that real world targets are easier to find in contextually consistent locations (Biederman, Mezzanotte, & Rabinowitz, 1982; Henderson et al., 1999), and prior weighting or the influence of global gist would explain this result. Search efficiency—our ability to avoid looking at oranges when the target is a pint of milk, or birds when we're looking for a car—is thus captured by top-down guidance in terms of both target features and probable locations.

How does saliency affect search? Research into this question has focussed on how bottom-up saliency is combined with top-down guidance. It has previously been suggested that search completely overrides saliency, either by bypassing the computation completely or by zero-weighting regions that are inconsistent with the task at hand (Underwood et al., 2006; Zelinsky et al., 2005). However, subsequent experiments in this thesis argue instead that saliency does have an influence in search, both when it coincides with the target and when it does not. For example in Experiment 4 salient regions were found more quickly and this was at least partly to do with their visual saliency. In Experiments 8 and 9, salient distractors slowed search more than non-salient ones, even when their relevance in terms of shared features with the target did not differ. Interestingly, the distractors in this case were located close to the target, and so their general direction may have been enhanced top-down. The effects of saliency in search

are consistent with research showing the capture of attention by salient distractors in simpler visual search (Nothdurft, 2002) and they justify the inclusion of the additive saliency component in my eye-guidance framework. This argues against research that suggests saliency does not play any role in search.

In Experiments 1(a), 6 and 7 there was little or no effect of saliency on search. How should this discrepancy be addressed? One possibility is that in these experiments the task and target information was more complete, meaning that there was more activation in the top-down relevance map so that saliency made less of a difference. However, when the information provided by the target template was manipulated (with a verbal label as opposed to a pictorial cue, or a category versus an instance target) there was no change in the effect of saliency. Searching for a picture was quicker than searching for an object by name, consistent with research by Wolfe et al. (2004) and Vickery et al. (2005). This more efficient top-down guidance did not lead to more or less of an effect of saliency.

An alternative way to resolve the different effects of saliency in search is to suggest that the contribution of saliency and relevance varies as a function of time or cognitive load. It was interesting that in a follow-up experiment to the search tasks from Experiment 1(a) some clear effects of target saliency were found when there was a time limit. The distractor effects in Experiment 9 were also found when there was a relatively fast trial time. Van Zoest et al. (2005) argue that bottom-up guidance occurs earlier than top-down guidance, and decays rapidly after about 250ms in a covert attention task. Other experiments in this thesis show that later fixations are still attracted to salient regions to some extent. If saliency reflects an earlier process than task- and knowledge-based guidance, then those search tasks without a time limit may allow top-down guidance to dominate. This would explain why effects were not always found in the present thesis.

A final point of discussion is how the framework outlined above can explain search behaviour when visual or semantic information is missing. The gaze-contingent experiments in Chapter 7 revealed some interesting effects. First, if the image was filtered so as to change the saliency map, the advantage participants showed in finding

certain regions changed (although for the most part the filtering used in the experiment did not achieve this). This is consistent with a saliency map which is computed from peripheral features and which contributes to guiding eye movements in search. Second, fixation durations were lengthened by this filtering. If there is a reciprocal relationship between the evolution of the different maps and the time to initiate a saccade, this delay may be explained by the saliency and relevance maps taking longer to accumulate due to the degraded information. When the peripheral background was fully masked, saliency and relevance will be largely unavailable, and so one should expect viewing to be driven largely by prior expectations. In the previous chapter I described a patient who, due to her agnosia, might not be able to set up a target template. This would result in a flat or reduced relevance map, and thus saccades would be targeted more on the basis of saliency. In the search task this was seen in the fact she looked at salient regions more often, and had higher saliency at fixation, than normal controls.

12.3 Concluding remarks and future directions

To conclude, the experiments in this thesis argue for the role of a saliency map within a larger framework of eye guidance in natural scenes. The case for the saliency map model should not be overstated, however. Within relatively unconstrained encoding tasks, raw bottom-up saliency was able to predict fixation locations better than chance, although estimates of its performance were somewhat less than those reported previously. This was particularly true when image-independent biases were taken into account, and these can potentially explain as much of where people look as can the saliency model. In order to make causal conclusions about the link between saliency and fixation an experimental approach is necessary. When the model was used to screen objects and regions it was useful in predicting those which were more or less likely to attract attention, and which were easier to find in some search situations. Within the context of a natural scene, the saliency map model can narrow down areas that are semantically informative. Indeed, one might speculate that what it is really doing is excluding parts of the image such as open sky and background where there are no abrupt changes in visual features that might indicate an object.

The framework I have outlined in this chapter is useful for suggesting future research on this topic. If saliency is indeed a separate process from top-down relevance then it might be dissociable in the time course of an experiment, or represented in different brain areas. The work in Chapter 11 could be taken as an indication that temporal and parietal areas influence eye guidance in scenes, and activation in these areas might contribute to a modulated saccade map like those identified in the frontal eye fields (Thompson, Bichot, & Sato, 2005). More precise modelling of exactly what features make up scene gist, and how this influences eye guidance, would be a useful avenue for further research.

There are several findings from this thesis that are not explained by the framework discussed. In Chapters 4, 5 and 8 I spent some time discussing the sequential patterns people make when moving their eyes and whether these scanpaths are consistent. The results are promising in that they suggest a systematic component that is thus far not explained by models of eye movement control. At its simplest, the findings regarding scanpaths suggest that some ways of moving around an image are more likely to be executed than others. Whether scanpaths are idiosyncratic, and the degree to which they reflect individuals' cognitions about complex scenes, are interesting topics to study further. The control of fixation duration and of within-object landing position are other unresolved issues, and some of the work presented here might be a starting point for investigating them further.

Part of the impetus for this thesis was the desire to move eye movement research closer to real stimuli, and it did so with mixed success. Using natural scenes allowed influences of context and semantic interpretation to be explored in a way that would not be possible with much simpler stimuli. In some search tasks it was more useful to use simpler object arrays in order to control for some of these influences. Of course, this research was limited by the laboratory set-up and some of the tasks were rather artificial. It would be interesting to see whether some of the effects found here would generalise to more realistic settings, with participants free to move around the environment, and with the added complexity of head and body movements. In particular, some of my observations regarding scene-centred biases might be different. The addition of motion to

the bottom-up model would be likely to make a big difference to the results (Itti, 2005), and this would be interesting to explore further.

Low-level, visual saliency does not explain or determine exactly where people look in their environment. Instead, the visual-cognitive system is largely efficient in selecting regions of interest in accordance with ongoing tasks. However, a relatively simple computational model of distinctiveness does tell us something about where people fixate, and this combines with the human interpretation of the scene. Rather than passively observing, humans interact with their world, and their knowledge interacts with the structure of the visual array to produce a scanpath. Saliency is one small part of this process.

References

Abed, F. (1991). Cultural Influences On Visual Scanning Patterns. *Journal Of Cross-Cultural Psychology, 22*(4), 525-534.

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal Of Experimental Psychology-Learning Memory And Cognition, 25*(4), 997-1010.

Barton, J. J. S., Behrmann, M., & Black, S. (1998). Ocular search during line bisection - The effects of hemi-neglect and hemianopia. *Brain, 121*, 1117-1131.

Barton, J. J. S., & Cherkasova, M. (2003). Face imagery and its relation to perception and covert recognition in prosopagnosia. *Neurology, 61*(2), 220-225.

Barton, J. J. S., Radcliffe, N., Cherkasova, M. V., & Edelman, J. A. (2007). Scan patterns during the processing of facial identity in prosopagnosia. *Experimental Brain Research, 181*(2), 199-211.

Becker, W. (1991). Saccades. In R. H. S. Carpenter (Ed.), *Eye movements* (pp. 95-137). London: Macmillan Press.

Becker, W., & Jurgens, R. (1990). Human Oblique Saccades - Quantitative-Analysis Of The Relation Between Horizontal And Vertical Components. *Vision Research, 30*(6), 893-920.

Behrmann, M. (2003). Neuropsychological approaches to perceptual organization. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects and scenes: Analytic and holistic processes* (pp. 295-334). Oxford: Oxford University Press.

Benson, D. F., Davis, R. J., & Snyder, B. D. (1988). Posterior Cortical Atrophy. *Archives Of Neurology, 45*(7), 789-793.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene Perception - Detecting And Judging Objects Undergoing Relational Violations. *Cognitive Psychology, 14*(2), 143-177.

Biederman, I., Rabinowitz, J. C., Glass, A., & Stacy, E. (1974). On the information extracted from a glance at a scene. *Journal Of Experimental Psychology-General, 103*, 597-560.

Brandt, H. F. (1945). *The psychology of seeing*. New York: The Philisophical Library.

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal Of Cognitive Neuroscience, 9*(1), 27-38.

Braun, J. (2003). Natural scenes upset the visual applecart. *Trends In Cognitive Sciences, 7*(1), 7-9.

Broadbent, D. E. (1954). The Role Of Auditory Localization In Attention And Memory Span. *Journal Of Experimental Psychology, 47*(3), 191-196.

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art.* Chicago: University of Chicago Press.

Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal Of Experimental Psychology-Human Perception And Performance, 33*(4), 753-763.

Chen, X., & Zelinsky, G. J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research, 46*(24), 4118-4133.

Choi, Y. S., Mosley, A. D., & Stark, L. (1995). String editing analysis of human visual search. *Optometry and vision science, 72*(7), 439-451.

Clavagnier, S., Berger, M. F., Klockgether, T., Moskau, S., & Karnath, H. O. (2006). Restricted ocular exploration does not seem to explain simultanagnosia. *Neuropsychologia, 44*(12), 2330-2336.

Collewijn, H., Erkelens, C. J., & Steinman, R. M. (1988). Binocular Coordination Of Human Vertical Saccadic Eye-Movements. *Journal Of Physiology-London, 404*, 183-197.

Coppola, D. M., Purves, H. R., McCoy, A. N., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings Of The National Academy Of Sciences Of The United States Of America, 95*(7), 4002-4006.

Crundall, D., & Underwood, G. (1998). Effects of experience and processing demands on visual information acquisition in drivers. *Ergonomics, 41*(4), 448-458.

Crundall, D., Underwood, G., & Chapman, P. (2002). Attending to the peripheral world while driving. *Applied Cognitive Psychology, 16*(4), 459-475.

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52*, 317-329.

Delvenne, J., Seron, X., Coyette, F., & Rossion, B. (2004). Evidence for perceptual deficits in associative visual (prosop)agnosia: a single-case study. *Neuropsychologia, 42*, 597-612.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review Of Neuroscience, 18*, 193-222.

Farah, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision*. Cambridge, MA: MIT Press.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active Vision: The Psychology of Looking and Seeing*. Oxford: OUP.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral And Brain Sciences, 22*(4), 661-721.

Frey, H. P., Konig, P., & Einhauser, W. (2007). The role of first- and second-order stimulus features for human overt attention. *Perception & Psychophysics, 69*(2), 153-161.

Friedman, A. (1979). Framing pictures: The role of knowledge in automised encoding and memory for gist. *Journal Of Experimental Psychology-General, 108*, 316-355.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5*(3), 490-495.

Frith, C. (2005). The top in top-down attention. In L. Itti, G. Rees & J. K. Tsotsos (Eds.), *Neurobiology of attention.* London: Elsevier.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673-1682.

Giersch, A., Humphreys, G. W., Boucart, M., & Kovacs, I. (2000). The computation of occluded contours in visual agnosia: Evidence for early computation prior to shape binding and figure-ground coding. *Cognitive Neuropsychology, 17*(8), 731-759.

Gilbert, C., Ito, M., Kapadia, M., & Westheimer, G. (2000). Interactions between attention, context and learning in primary visual cortex. *Vision Research, 40*, 1217-1226.

Gilchrist, I. D., Brown, V., & Findlay, J. M. (1997). Saccades without eye movements. *Nature, 390*(6656), 130-131.

Gilchrist, I. D., & Harvey, M. (2006). Evidence for a systematic component within scan paths in visual search. *Visual Cognition, 14*(4-8), 704-715.

Gottlieb, J. (2002). Parietal mechanisms of target representation. *Current Opinion In Neurobiology, 12*(2), 134-140.

Greene, H. H. (2006). The control of fixation duration in visual search. *Perception, 35*(3), 303-315.

Groner, R., Walder, F., & Groner, M. (1984). Looking at faces: local and global aspects of scanpaths. In A. Gale & F. Johnson (Eds.), *Theoretical and applied aspects of eye movement research* (pp. 523-533). Amsterdam: Elsevier.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends In Cognitive Sciences, 9*(4), 188-194.

Henderson, J. M. (1993). Eye-Movement Control During Visual Object Processing - Effects Of Initial Fixation Position And Semantic Constraint. *Canadian Journal Of Experimental Psychology, 47*(1), 79-98.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends In Cognitive Sciences, 7*(11), 498-504.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. L. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. van Gompel, M. Fischer, W. Murray & R. W. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537-562). Amsterdam: Elsevier.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal Of Experimental Psychology-Human Perception And Performance, 25*(1), 210-228.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition, 33*(1), 98-106.

Hillstrom, A. P., & Yantis, S. (1994). Visual-Motion And Attentional Capture. *Perception & Psychophysics, 55*(4), 399-411.

Hughes, A. (1977). The topography of vision in mammals of contrasting lifestyles. In F. Crescitelli (Ed.), *Handbook of sensory physiology* (Vol. VII, pp. 614-642). Berlin: Springer.

Hugli, H., Jost, T., & Ouerhani, N. (2005). Model performance for visual attention in real 3D color scenes. In *Artificial Intelligence And Knowledge Engineering Applications: A Bioinspired Approach, Pt 2, Proceedings* (Vol. 3562, pp. 469-478).

Humphreys, G. W., & Riddoch, M. J. (1987). *To see but not to see: A case study of visual agnosia*. Hillsdale, NJ: Erlbaum.

Husk, J. S., Bennett, P. J., & Sekuler, A. B. (2007). Inverting houses and textures: Investigating the characteristics of learned inversion effects. *Vision Research, 47*(9), 3350-3359.

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition, 12*(6), 1093-1123.

Itti, L. (2006). Quantitative modelling of perceptual salience at human eye position. *Visual Cognition, 14*(4-8), 959-984.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489-1506.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194-203.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *Ieee Transactions On Pattern Analysis And Machine Intelligence, 20*(11), 1254-1259.

Jovancevic, J., Sullivan, B., & Hayhoe, M. (2006). Control of attention and gaze in complex environments. *Journal Of Vision, 6*(12), 1431-1450.

Kelley, T. A., Chun, M. M., & Chua, K. P. (2003). Effects of scene inversion on change detection of targets matched for visual salience. *Journal Of Vision, 3*(1), 1-5.

Klein, R. M. (2000). Inhibition of return. *Trends In Cognitive Sciences, 4*(4), 138-147.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology, 4*, 219-227.

Kramer, A. F., Hahn, S., Irwin, D. E., & Theeuwes, J. (1999). Attentional capture and aging: Implications for visual search performance and oculomotor control. *Psychology And Aging, 14*(1), 135-154.

Laeng, B., & Teodorescu, D. S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science, 26*(2), 207-231.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*(11), 1311-1328.

Leonards, U., & Scott-Samuel, N. E. (2005). Idiosyncratic initiation of saccadic face exploration in humans. *Vision Research, 45*(20), 2677-2684.

Li, Z. P. (2002). A saliency map in primary visual cortex. *Trends In Cognitive Sciences, 6*(1), 9-16.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive Determinants Of Fixation Location During Picture Viewing. *Journal Of Experimental Psychology-Human Perception And Performance, 4*(4), 565-572.

Loschky, L. C., & McConkie, G. W. (2002). Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display. *Journal Of Experimental Psychology-Applied, 8*(2), 99-117.

Loschky, L. C., McConkie, G. W., Yang, H., & Miller, M. E. (2005). The limits of visual resolution in natural scene viewing. *Visual Cognition, 12*(6), 1057-1092.

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics, 2*, 547-552.

Mannan, S., Ruddock, K., & Wooding, D. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision, 9*(3), 363-386.

Mannan, S., Ruddock, K., & Wooding, D. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision, 10*(10), 65-188.

Mannan, S., Ruddock, K., & Wooding, D. (1997). Fixation patterns made during brief examination of two dimensional images. *Perception, 26*, 1059-1072.

Martin, T., Riley, M. E., Kelly, K. N., Hayhoe, M., & Huxlin, K. R. (2007). Visually-guided behavior of homonymous hemianopes in a naturalistic task. *Vision Research, 47*(28), 3434-3446.

McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading I: the location of initial eye fixations on words. *Vision Research, 28*(10), 1107-1118.

McConkie, G. W., & Rayner, K. (1975). Span Of Effective Stimulus During A Fixation In Reading. *Perception & Psychophysics, 17*(6), 578-586.

McPhee, L. C., Scialfa, C. T., Dennis, W. M., Ho, G., & Caird, J. K. (2004). Age differences in visual search for traffic signs during a simulated conversation. *Human Factors, 46*(4), 674-685.

Morrison, R. E. (1984). Manipulation Of Stimulus Onset Delay In Reading - Evidence For Parallel Programming Of Saccades. *Journal Of Experimental Psychology-Human Perception And Performance, 10*(5), 667-682.

Munoz, D. P., Broughton, J. R., Goldring, J. E., & Armstrong, I. T. (1998). Age-related performance of human subjects on saccadic eye movement tasks. *Experimental Brain Research, 121*(4), 391-400.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45*(2), 205-231.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46*(5), 614-621.

Neider, M. B., & Zelinsky, G. J. (2008). Exploring set size effects in scenes: Identifying the objects of search. *Visual Cognition, 16*(1), 1-10.

Nothdurft, H. C. (2002). Attention shifts to salient targets. *Vision Research, 42*(10), 1287-1306.

Noton, D., & Stark, L. (1971). Scanpaths In Saccadic Eye Movements While Viewing And Recognizing Patterns. *Vision Research, 11*(9), 929-&.

O'Regan, J. K., & Jacobs, A. M. (1992). Optimal Viewing Position Effect In Word Recognition - A Challenge To Current Theory. *Journal Of Experimental Psychology-Human Perception And Performance, 18*(1), 185-197.

O'Regan, J. K., Levy-Schoen, A., Pynte, J., & Brugaillere, B. (1984). Convenient Fixation Location Within Isolated Words Of Different Length And Structure. *Journal Of Experimental Psychology-Human Perception And Performance, 10*(2), 250-257.

Ohman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal Of Experimental Psychology-General, 130*(3), 466-478.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends In Cognitive Sciences, 11*(12), 520-527.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*(1), 107-123.

Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision, 16*(2), 125-154.

Pelz, J., Hayhoe, M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research, 139*(3), 266-277.

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research, 45*(18), 2397-2416.

Pomplun, M. (2006). Saccadic selectivity in complex visual search displays. *Vision Research, 12*(46), 1886.

Pomplun, M., Reingold, E. M., & Shen, J. Y. (2001). Peripheral and parafoveal cueing and masking effects on saccadic selectivity in a gaze-contingent window paradigm. *Vision Research, 41*(21), 2757-2769.

Posner, M. I. (1980). Orienting Of Attention. *Quarterly Journal Of Experimental Psychology, 32*, 3-25.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal Of Experimental Psychology-Learning Memory And Cognition, 2*, 509-522.

Privitera, C. M. (2006). The scanpath theory: its definition and later developments. *Proceedings of SPIE (The international Society for Optical Engineering), 6057*.

Privitera, C. M., & Stark, L. W. (2000). Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions On Pattern Analysis And Machine Intelligence, 22*(9), 970-982.

Rabbitt, P. (1965). An age-decrement in the ability to ignore irrelevant information. *Journal of Gerontology, 20*, 233-237.

Raj, R., Geisler, W. S., Frazor, R. A., & Bovik, A. C. (2005). Contrast statistics for foveated visual systems: fixation selection by minimizing contrast entropy. *Journal Of The Optical Society Of America A-Optics Image Science And Vision, 22*(10), 2039-2049.

Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research, 42*(11), 1447-1463.

Rayner, K. (1979). Eye Guidance In Reading - Fixation Locations Within Words. *Perception, 8*(1), 21-30.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network-Computation In Neural Systems, 10*(4), 341-350.

Reingold, E. M., Loschky, L. C., McConkie, G. W., & Stampe, D. M. (2003). Gaze-contingent multiresolutional displays: An integrative review. *Human Factors, 45*(2), 307-328.

Renninger, L. W., Verghese, P., & Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. *Journal Of Vision, 7*(3).

Riddoch, M. J., & Humphreys, G. W. (1987). A Case Of Integrative Visual Agnosia. *Brain, 110*, 1431-1462.

Rizzo, M., & Hurtig, R. (1987). Looking But Not Seeing - Attention, Perception, And Eye-Movements In Simultanagnosia. *Neurology, 37*(10), 1642-1648.

Rizzolatti, G., Riggio, L., Dascola, I., & Umilta, C. (1987). Reorienting Attention Across The Horizontal And Vertical Meridians - Evidence In Favor Of A Premotor Theory Of Attention. *Neuropsychologia, 25*(1A), 31-40.

Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychological Science, 11*(6), 454-461.

Sankhoff, D., & Kruskal, J. B. (Eds.). (1983). *Time warps, string edits and macromolecules: The theory and practice of sequence comparison*. Reading, MA: Addison-Wesley.

Sanocki, T. (2003). Representation and perception of scenic layout. *Cognitive Psychology, 47*(1), 43-86.

Schmidt, J., & Zelinsky, G. (2007). Manipulating the availability of visual information in search, *Vision Sciences Society*. Sarasota, Florida.

Scialfa, C. T., Hamaluk, E., Skaloud, P., & Pratt, J. (1999). Age differences in saccadic averaging. *Psychology And Aging, 14*, 695-699.

Scialfa, C. T., & Joffe, K. M. (1997). Age differences in feature and conjunction search: Implications for theories of visual search and generalized slowing. *Aging Neuropsychology And Cognition, 4*(3), 227-246.

Scialfa, C. T., Thomas, D. M., & Joffe, K. M. (1994). Age-Differences In The Useful Field-Of-View - An Eye-Movement Analysis. *Optometry And Vision Science, 71*(12), 736-742.

Shimozaki, S. S., Hayhoe, M. M., Zelinsky, G. J., Weinstein, A., Merigan, W. H., & Ballard, D. H. (2003). Effect of parietal lobe lesions on saccade targeting and spatial memory in a naturalistic visual search task. *Neuropsychologia, 41*(10), 1365-1386.

Smeets, J. B. J., Hayhoe, M. M., & Ballard, D. H. (1996). Goal-directed arm movements change eye-head coordination. *Experimental Brain Research, 109*(3), 434-440.

Stark, L., & Ellis, S. R. (1981). Scanpaths revisited: cognitive models direct active looking. In D. F. Fisher, R. A. Monty & J. W. Senders (Eds.), *Eye movements: cognition and visual perception* (pp. 193-227). Hillsdale, NJ: Lawrence Erlbaum.

Stark, L., & Noton, D. (1971). Scanpaths And Pattern Recognition. *Science, 173*(3998), 753-&.

Stirk, J. A., & Underwood, G. (2007). Low-level visual saliency does not predict change detection in natural scenes. *Journal of Vision, 7*(10), 1-10.

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research, 45*(5), 643-659.

Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research, 46*(12), 1857-1862.

Thompson, K. G., Bichot, N. P., & Sato, T. R. (2005). Frontal eye field activity before visual search errors reveals the integration of bottom-up and top-down salience. *Journal Of Neurophysiology, 93*(1), 337-351.

Toet, A., Bijl, P., & Valeton, J. M. (2001). Image dataset for testing search and detection models. *Optical Engineering, 40*(9), 1760-1767.

Torralba, A. (2003). Modeling global scene factors in attention. *Journal Of The Optical Society Of America A-Optics Image Science And Vision, 20*(7), 1407-1418.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113*(4), 766-786.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Treisman, A., & Gormican, S. (1988). Feature Analysis In Early Vision - Evidence From Search Asymmetries. *Psychological Review, 95*(1), 15-48.

Treue, S. (2003). Visual attention: the where, what, how and why of saliency. *Current Opinion In Neurobiology, 13*(4), 428-432.

Turano, K. A., Geruschat, D. R., & Baker, F. H. (2003). Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research, 43*(3), 333-346.

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal Of Experimental Psychology, 59*(11), 1931-1949.

Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal Of Cognitive Psychology, 18*(3), 321-342.

Valentine, T. (1988). Upside-Down Faces - A Review Of The Effect Of Inversion Upon Face Recognition. *British Journal Of Psychology, 79*, 471-491.

van Diepen, P. M. J., & d'Ydewalle, G. (2003). Early peripheral and foveal processing in fixations during scene perception. *Visual Cognition, 10*(1), 79-100.

van Zoest, W., & Donk, M. (2005). The effects of salience on saccadic target selection. *Visual Cognition, 12*(2), 353-375.

van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal Of Experimental Psychology-Human Perception And Performance, 30*(4), 746-759.

Vickery, T. J., King, L. W., & Jiang, Y. H. (2005). Setting up the target template in visual search. *Journal Of Vision, 5*(1), 81-92.

Vincent, B. T., Troscianko, T., & Gilchrist, I. D. (2007). Investigating a space-variant weighted salience account of visual selection. *Vision Research, 47*(13), 1809-1820.

Walker, R., Deubel, H., Schneider, W. X., & Findlay, J. M. (1997). Effect of remote distractors on saccade programming: Evidence for an extended fixation zone. *Journal Of Neurophysiology, 78*(2), 1108-1119.

Williams, C. C., Henderson, J. M., & Zacks, R. T. (2005). Incidental visual memory for targets and distractors in visual search. *Perception and Psychophysics, 67*(5), 816-827.

Wolfe, J. M. (1994). Guided Search 2.0 - A Revised Model Of Visual-Search. *Psychonomic Bulletin & Review, 1*(2), 202-238.

Wolfe, J. M. (1998a). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13-71). Hove: Psychology Press.

Wolfe, J. M. (1998b). What can 1 million trials tell us about visual search? *Psychological Science, 9*(1), 33-39.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research, 44*(12), 1411-1426.

Wooding, D. (2002). Eye movements of large populations: II. Deriving regions of interest, coverage and similarity using fixation maps. *Behavior Research Methods Instruments & Computers, 34*(4), 518-528.

Wright, M. J. (2005). Saliency predicts change detection in pictures of natural scenes. *Spatial Vision, 18*(4), 413-430.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum.

Yee, R. D., Schiller, V. L., Lim, V., Baloh, F. G., Baloh, R. W., & Honrubia, V. (1985). Velocities Of Vertical Saccades With Different Eye-Movement Recording Methods. *Investigative Ophthalmology & Visual Science, 26*(7), 938-944.

Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel-serial visual search. *Journal Of Experimental Psychology-Human Perception And Performance, 23*(1), 244-262.

Zelinsky, G. J., Zhang, W., Yu, B., Chen, X., & Samaras, D. (2005). *The role of top-down and bottom-up processes in guiding eye movements in visual search.* Paper presented at the Neural Information Processing Systems.

Zetzsche, C. (2005). Natural scene statistics and salient visual features. In L. Itti, G. Rees & J. K. Tsotsos (Eds.), *Neurobiology of Attention*. London: Elsevier.

List of Appendices

Appendix A: Scanpath comparison methods

Java code for string edit distance

```
public double stringEditDistance()
{

//*****************************
// Compute Levenshtein distance
//*****************************
//based on code at http://www.merriampark.com/ld.htm

    int d[ ][ ]; // matrix
    String a; // string from scanpath a
    String b; // string from scanpath b
    int n; // length of scanpath a
    int m; // length of scanpath b
    int i; // iterates through a
    int j; // iterates through b
    char a_i; // ith character of a
    char b_j; // jth character of b
    int cost; // cost

    double sd; // the edit distance
    double sdNorm; // the normalized edit distance
    double sdSim; // the normalized similarity

    // if either string is empty, distance is length of the
other
    n = a.length ();
    m = b.length ();
    if (n == 0)
    {
        sd = m;
        return sd;
    }
    if (m == 0)
    {
        sd = n;
        return sd;
    }

    d = new int[n+1][m+1];

    // set the first row/column to integers ascending from 1
    for (i = 0; i <= n; i++)
    {
        d[i][0] = i;
    }
```

```
    for (j = 0; j <= m; j++)
    {
        d[0][j] = j;
    }

    // loop through the first string
    for (i = 1; i <= n; i++)
    {
        s_i = a.charAt (i - 1);

        // loop through the second string
        for (j = 1; j <= m; j++)
        {
                t_j = b.charAt (j - 1);

                // compare the two characters
                if (s_i == t_j)
                {
                cost = 0;
                }
                else
                {
                cost = 1;       //cost for unequal
characters is 1
                }

        // set the current cell
        d[i][j]
        = findlowest (d[i-1][j]+1, d[i][j-1]+1, d[i-1][j-1]
+ cost);
        }
    }

    // lowest distance is bottom left cell
    sd = d[n][m];

    //normalise the distance over the length of the longer
string
    if (n>=m)
    {
        sdNorm = sd/n;
    }
    else
    {
        sdNorm = sd/m;
    }

//similarity is 1 minus the normalised distance
sdSim = 1-sdNorm;
return sd;
}
```

```
public int findlowest(int int1,int int2,int int3)
//method to find the lowest of three integers
{
 if (int2<int1)
 {
    int1=int2;
 }
 if (int3<int1)
 {
    int1=int3;
 }
    return int1;
}
```

Java code for Mannan linear distance

```
public double mannanDistance(int x, int y)
 //****************************
 // Compute Mannan distance
 //****************************
 //based on Mannan et al (1995)

{
        double d[ ][ ]; // matrix
        Point spA[ ]; // first scanpath (an array of points)
        Point spB[ ]; // second scanpath
        int n; // length of first
        int m; // length of second
        int i; // iterates through first
        int j; // iterates through second
        double d1i = 0;
        //the sum of squared distances from the ith fixation
        in the 1st scanpath to its nearest neighbour
        double d2j = 0;
        //the sum of squared distances from the jth fixation
      in the 2nd scanpath to its nearest neighbour
        double md; // the normalized, mean linear distance

//make a matrix of the distance between each of the
fixations
//the lowest in each column/row gives the nearest neighbour
distance

        d=new double [n][m];
        for(i=0;i<n;i++)
        {
                for(j=0;j<m;j++)
                {
                //get the distance
```

```
                d[i][j] = spA[i].distance( spB[j] );
                }
        }

        double lowest;
        for(i=0;i<n;i++)
        {
        lowest=100000;

                for(j=0;j<m;j++)
                {
                        if (d[i][j]<lowest)
                        {
                                lowest=d[i][j];
                        }
                }

                //sum the squared distances from A to B
                d1i=d1i+(lowest*lowest);

        }

        for(j=0;j<m;j++)
        {
        lowest=100000;
                for(i=0;i<n;i++)
                {
                        if (d[i][j]<lowest)
                        {
                                lowest=d[i][j];
                        }
                }

                //sum the squared distances from B to A
                d2j=d2j+(lowest*lowest);
        }

//multiply the sum of squares by the length of each scanpath
        double dsquared = (n*d2j)+(m*d1i);
        //normalise over the display size
        dsquared = dsquared / ((2*n*m)*((x*x)+(y*y)));
        md = Math.sqrt(dsquared);
        return md;
        }
```

Appendix B: Formula for 2D correlation

$$r = \frac{\sum_m \sum_n \left(A_{mn} - \overline{A}\right)\left(B_{mn} - \overline{B}\right)}{\sqrt{\left(\sum_m \sum_n \left(A_{mn} - \overline{A}\right)^2\right)\left(\sum_m \sum_n \left(B_{mn} - \overline{B}\right)^2\right)}}$$

where r is the correlation between two matrices, A and B which have the same dimensions, m and n.

Appendix C: Edge orientation operators

Edges can be detected by convolution with Sobel kernels

$$G_x = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix}$$

$$G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

where *Gx* and *Gy* give the local gradient in the horizontal and vertical directions respectively.

The absolute edge magnitude, *G*, is given by

$$G = \sqrt{G_x{}^2 + G_y{}^2}$$

The direction, θ, of each pixel is given by

$$\theta = \arctan\left(\frac{G_y}{G_x}\right)$$

and the summed magnitude in this direction is

$$\sum_{p=1,\ldots,n} G_p$$

where *n* is the number of pixels at that particular orientation and G_p is the edge magnitude at each point.

Appendix D: Statistical data

Experiment 1(a)

Mixed ANOVA on ordinal fixation on target

Source	***SS***	***df***	***MS***	***F***	***p***
Task	2462.058	2	1231.029	67.937	.000
Error (Task)	761.049	42	18.120		
Saliency	40.229	1	40.229	13.564	.001
Error (Saliency)	124.563	42	2.966		
Task x Saliency	28.522	2	14.261	4.809	.013

Post hoc tests on levels of task

Comparison	**Statistic**
Memory v. category	102.34****
Memory v. instance	101.54****
Category v. instance	0.000

Simple main effects

Source	***SS***	***df***	***MS***	***F***	***p***
Task at low saliency	1509.329	2	754.664	71.604	0.000
Task at medium saliency	981.271	2	490.636	46.552	0.000
Error	885.316	84	10.539		
Saliency at memory	64.153	1	64.153	21.634	0.000
Saliency at category	3.228	1	3.228	1.088	0.3028
Saliency at instance	1.404	1	1.404	0.473	0.4952
Error	124.543	42	2.965		

Mixed ANOVA on probability of target fixation

Source	***SS***	***df***	***MS***	***F***	***p***
Task	0.260	2	0.130	2.069	.139
Error (Task)	2.643	42	0.063		
Saliency	0.015	1	0.015	3.777	.059
Error (Saliency)	0.172	42	0.004		
Task x Saliency	0.060	2	0.030	7.321	.002

Simple main effects

Source	***SS***	***df***	***MS***	***F***	***p***
Task at low saliency	0.272	2	0.136	4.047	0.021
Task at medium saliency	0.045	2	0.023	0.674	0.5124
Error	2.822	84	0.034		
Saliency at memory	0.065	1	0.065	15.997	0.0003
Saliency at category	0.002	1	0.002	0.47	0.4967
Saliency at instance	0.007	1	0.007	1.803	0.1866
Error	0.172	42	0.004		

Mixed ANOVA on first gaze duration on target

Source	***SS***	***df***	***MS***	***F***	***p***
Task	549484.613	2	274742.306	3.578	0.037
Error (Task)	3225156.123	42	76789.432		
Saliency	31877.387	1	31877.387	2.497	0.122
Error (Saliency)	536207.905	42	12766.855		
Task x Saliency	25277.916	2	12638.958	0.99	0.38

Post hoc tests on levels of task

Comparison	**Statistic**
Memory v. category	7.04*
Memory v. instance	1.06
Category v. instance	2.63

Mixed ANOVA on total picture inspection duration

Source	***SS***	***df***	***MS***	***F***	***p***
Task	1219965506	2	609982752.9	43.144	.000
Error (Task)	593809199.5	42	14138314.27		
Saliency	316355.711	1	316355.711	0.989	.326
Error (Saliency)	13437463.95	42	319939.618		
Task x Saliency	25860.548	2	12930.274	0.040	.960

Post hoc tests on levels of task

Comparison	**Statistic**
Memory v. category	66.14****
Memory v. instance	63.26****
Category v. instance	0.03

Mixed ANOVA on number of target fixations

Source	***SS***	***df***	***MS***	***F***	***p***
Task	16.875	2	8.437	7.535	.002
Error (Task)	47.027	42	1.120		
Saliency	1.206	1	1.206	8.248	.006
Error (Saliency)	6.138	42	.146	6.138	42
Task x Saliency	.599	2	.300	2.050	.141

Post hoc tests on levels of task

Comparison	**Statistic**
Memory v. category	11.39**
Memory v. instance	11.2**
Category v. instance	0

Mixed ANOVA on proportion of correct responses

Source	***SS***	***df***	***MS***	***F***	***p***
Task	0.009	1	0.009	0.293	0.593
Error (Task)	0.884	28	0.032		
Saliency	0.028	1	0.028	1.16	0.291
Error (Saliency)	0.683	28	0.024		
Task x Saliency	0.002	1	0.002	0.084	0.773

Mixed ANOVA on probability of fixating the most salient region

Source	***SS***	***df***	***MS***	***F***	***p***
Task	8.210	2	4.105	191.24	0.000
Error (Task)	0.902	42	0.021		
Saliency	0.057	1	0.057	6.309	0.016
Error (Saliency)	0.382	42	0.009		
Task x Saliency	0.019	2	0.009	1.022	0.369

Post hoc tests on levels of task

Comparison	**Statistic**
Memory v. category	298.7****
Memory v. instance	281.8****
Category v. instance	0.25

Mixed ANOVA on probability of fixating the most salient region (first 5 fixations only)

Source	***SS***	***df***	***MS***	***F***	***p***
Task	0.151	2	0.0756	3.995	0.026
Error (Task)	0.795	42	0.0189		
Saliency	0.0654	1	0.0654	5.018	0.030
Error (Saliency)	0.548	42	0.0130		
Task x Saliency	0.0129	2	0.0064	0.493	0.614

Post hoc tests on levels of task

Comparison	**Statistic**
Memory v. category	7.12*
Memory v. instance	4.59*
Category v. instance	0.28

Experiment 1(b)

Between-groups ANOVA on hit rate

Source	***SS***	***df***	***MS***	***F***	***p***
Group	0.250	1	0.250	18.104	0.000
Error (Group)	0.386	28	0.014		

Within-subjects ANOVA on hit rate

Source	***SS***	***df***	***MS***	***F***	***p***
Saliency	0.045	1	0.045	2.154	0.164
Error (Saliency)	0.295	14	0.021		

Within-subjects ANOVA on time to first fixate target

Source	***SS***	***df***	***MS***	***F***	***p***
Saliency	4055.963	1	4055.963	3.058	0.102
Error (Saliency)	18569.440	14	1326.389		

Within-subjects ANOVA on distance from first saccade landing site

Source	***SS***	***df***	***MS***	***F***	***p***
Saliency	69.669	1	69.669	15.496	0.001
Error (Saliency)	62.942	14	4.496		

Between-groups ANOVA on probability of fixating the most salient region

Source	***SS***	***df***	***MS***	***F***	***p***
Group	0.013	1	0.013	2.107	0.158
Error (Group)	0.174	28	0.006		

Within-subjects ANOVA on probability of fixating the most salient region

Source	***SS***	***df***	***MS***	***F***	***p***
Saliency	0.040	1	0.040	3.25	0.093
Error (Saliency)	0.174	14	0.012		

Experiment 2

Within-subjects ANOVA on number of fixations

Source	***SS***	***df***	***MS***	***F***	***p***
Task phase	1.030	1.217	0.847	2.323	0.136
Error (Task phase)	8.868	24.331	0.364		

NB. Greenhouse-Geisser corrected

Within-subjects ANOVA on fixation duration

Source	***SS***	***df***	***MS***	***F***	***p***
Task phase	3022.370	1.31	2308.012	3.653	0.057
Error (Task phase)	16545.584	26.19	631.746		

NB. Greenhouse-Geisser corrected

Within-subjects ANOVA on proportion of fixations on salient regions

Source	***SS***	***df***	***MS***	***F***	***p***
Task phase	0.023	2	0.012	75.604	0.000
Error (Task phase)	0.006	40	0.0002		

Post hoc paired samples *t* tests between levels of task phase

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Enc v. New	0.0015	0.01709	0.409	20	1.000
Enc v. Old	-0.0401	0.02108	-8.708	20	0.000
Old v. New	-0.0416	0.01379	-13.817	20	0.000

NB. *p* value is Bonferroni corrected.

Independent samples *t* tests between observed data and models

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Enc v. Random	0.067	0.00621	10.794	40	0.000
Enc v. Biased	0.0444	0.00622	7.127	40	0.000
Enc v. Trans	0.0497	0.00616	8.073	40	0.000
New v. Random	0.0655	0.0045	14.557	40	0.000
New v. Biased	0.0428	0.00452	9.472	40	0.000
New v. Trans	0.0482	0.00443	10.878	40	0.000
Old v. Random	0.107	0.00636	16.839	40	0.000
Old v. Biased	0.0844	0.00637	13.242	40	0.000
Old v. Trans	0.0898	0.00631	14.227	40	0.000

Within-subjects ANOVA on saliency value at fixation

Source	***SS***	***df***	***MS***	***F***	***p***
Task phase	73.895	2	36.948	2.697	0.080
Error (Task phase)	547.913	40	13.6980		

Independent samples *t* tests between observed data and models

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
v. Random	31.6045	0.81472	38.792	40	0.000
v. Biased	17.0676	0.70947	24.057	40	0.000
v. Trans	17.3446	1.69021	10.262	40	0.000

Within-subjects ANOVA on saliency value at fixation

Source	***SS***	***df***	***MS***	***F***	***p***
Task phase	369.708	2	184.854	2.699	0.080
Error (Task phase)	2739.771	40	68.4940		
Fixation number	1717.732	4	429.433	10.259	0.000
Error (Fix. num.)	3348.649	80	41.858		
Task phase x Fix. num.	146.115	8	18.264	0.514	0.844
Error (Task phase x Fix. num.)	5680.766	160	35.505		

Post hoc paired samples *t* tests between levels of fixation number

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
1 v. 2	-2.5099	4.42061	-2.602	20	.176
1 v. 3	2.2093	6.07999	1.665	20	1.000
1 v. 4	3.4053	6.32311	2.468	20	.196
1 v. 5	3.5088	5.2051	3.089	20	.043
2 v. 3	4.7191	4.89138	4.421	20	.003
2 v. 4	5.9151	5.36941	5.048	20	.001
2 v. 5	6.0187	5.63176	4.897	20	.001
3 v. 4	1.196	4.5167	1.213	20	1.000
3 v. 5	1.2995	5.78738	1.029	20	1.000
4 v. 5	0.1035	4.13564	0.115	20	1.000

NB. *p* value is Bonferroni corrected.

Within-subjects ANOVA on string-edit similarity

Source	***SS***	***df***	***MS***	***F***	***p***
Comparison	0.2850	2	0.1430	148.111	0.000
Error (Comparison)	0.0385	40	0.0010		

Post hoc paired samples *t* tests between levels of Comparison

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Enc-Old v. Old-New	0.1321	0.0461	13.136	20	0.000
Enc-Old v. Enc-New	0.1514	0.04542	15.271	20	0.000
Old-New v. Enc-New	0.0192	0.03985	2.21	20	0.117

NB. *p* value is Bonferroni corrected.

One-sample *t* tests on string edit similarity

Comparison	***M difference***	***t***	***df***	***p***
Enc-Old	0.2317	25.301	20	0.000
Old-New	0.0995	14.119	20	0.000
Enc-New	0.0803	10.587	20	0.000

Test value=0.0417

Within-subjects ANOVA on Mannan distances

Source	***SS***	***df***	***MS***	***F***	***p***
Comparison	6923.2190	2	3461.6090	185.649	0.000
Error (Comparison)	745.8390	40	18.6460		

Post hoc paired samples *t* tests between levels of Comparison

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Enc-Old v. Old-New	18.8854	7.43049	11.647	20	0.000
Enc-Old v. Enc-New	24.51	4.8703	23.062	20	0.000
Old-New v. Enc-New	5.6246	5.73967	4.491	20	0.001

NB. *p* value is Bonferroni corrected.

One-sample *t* tests on string edit similarity

Comparison	***M difference***	***t***	***df***	***p***
Enc-Old	0.2317	25.301	20	0.000
Old-New	0.0995	14.119	20	0.000
Enc-New	0.0803	10.587	20	0.000

Test value=0.0417

Within-subjects ANOVA on string edit comparison with saliency

Source	***SS***	***df***	***MS***	***F***	***p***
Task phase	0.0003	2	0.0001	0.539	0.587
Error (Task phase)	0.0107	40	0.0003		

Within-subjects ANOVA on Mannan comparison with saliency

Source	***SS***	***df***	***MS***	***F***	***p***
Task phase	187.0300	2	93.5150	15.867	0.000
Error (Task phase)	235.7470	40	5.8940		

Post hoc paired samples *t* tests between levels of task phase

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Enc v. New	-4.1814	4.1539	-4.613	20	0.001
Enc v. Old	-1.5947	3.33527	-2.191	20	0.121
Old v. New	-2.5867	2.64256	-4.486	20	0.001

NB. *p* value is Bonferroni corrected.

Experiment 3

Paired samples *t* test on region eccentricity

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	-0.1468	0.74277	-0.765	14	0.457
S1 v. Control	1.6766	1.07925	6.017	14	0.000
S5 v. Control	1.8234	1.49708	4.717	14	0.000

Within-subjects ANOVA on proportion of region fixations

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	0.1740	2	0.0868	31.773	0.000
Error (Region type)	0.0929	34	0.0027		

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	0.008	0.0661	0.515	17	1.000
S1 v. Control	0.1241	0.06841	7.694	17	0.000
S5 v. Control	0.116	0.08566	5.748	17	0.000

NB. *p* value is Bonferroni corrected.

Within-subjects ANOVA on time to first fixation

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	58383.2270	2	29191.6130	4.45	0.019
Error (Region type)	223047.0600	34	6560.2080		

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	13.9713	99.33175	0.597	17	1.000
S1 v. Control	-61.7084	119.04608	-2.199	17	0.126
S5 v. Control	-75.6797	123.78401	-2.594	17	0.057

NB. *p* value is Bonferroni corrected.

Within-subjects ANOVA on hit rate

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	0.0212	2	0.0106	1.471	0.244
Error (Region type)	0.2450	34	0.0072		

Within-subjects ANOVA on d prime

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	0.9060	2	0.4530	0.967	0.391
Error (Region type)	15.9340	34	0.4690		

Within-subjects ANOVA on RT

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	80008.8300	2	40004.4150	1.222	0.307
Error (Region type)	1112789.6100	34	32729.1060		

Experiment 4

Within-subjects ANOVA on time to first fixation (all trials)

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	115694.5	2	57847.3	1.287	0.291
Error (Region type)	1348303.3	30	44943.4		

Within-subjects ANOVA on time to first fixation (when cued)

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	68938.8	2	34469.4	0.818	0.451
Error (Region type)	1263417.4	30	42113.9		

Within-subjects ANOVA on probability of fixation in target absent trials

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	0.278	2	0.139	36.413	0.000
Error (Region type)	0.114	30	0.004		

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	0.0088	0.07205	0.488	15	1.000
S1 v. Control	0.1655	0.10174	6.508	15	0.000
S5 v. Control	0.1567	0.08559	7.325	15	0.000

NB. *p* value is Bonferroni corrected.

Within-subjects ANOVA on hit rate

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	0.018	2	0.009	1.096	0.347
Error (Region type)	0.246	30	0.008		

Within-subjects ANOVA on d prime

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	2.041	2	1.021	1.22	0.309
Error (Region type)	25.091	30	0.836		

Within-subjects ANOVA on RT

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	3191707.1	1.26	2526869.9	10.645	0.002
Error (Region type)	4497672.1	18.95	237386.7		

NB. Greenhouse-Geisser corrected

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	-509.4691	695.03232	-2.932	15	0.031
S1 v. Control	-578.0818	572.03847	-4.042	15	0.003
S5 v. Control	-68.6128	298.72477	-0.919	15	1.000

NB. *p* value is Bonferroni corrected.

Experiment 5(a)

Mixed ANOVA on time to fixate the target

Source	***SS***	***df***	***MS***	***F***	***p***
Filtering type	2419660.234	2	1209830.117	4.065	0.023
Error (Filtering)	14582048.83	49	297592.833		
Saliency	570641.134	2	285320.567	8.115	0.001
Error (Saliency)	3445793.79	98	35161.161		
Filtering x Saliency	38715.461	4	9678.865	0.275	0.893

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	-145.2893	256.1055	-4.091	51	0.001
S1 v. Control	-97.9494	236.25547	-2.99	51	0.014
S5 v. Control	47.3399	289.07518	1.181	51	0.772

NB. *p* value is Bonferroni corrected.

Post hoc independent samples *t* test between levels of filtering type

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Blur v. grey	-21.5663	116.76033	-0.185	33	1.000
Blur v. contrast	253.7357	97.60754	2.6	32	0.069
Grey v. contrast	275.302	105.26014	2.615	33	0.038

NB. *p* value is Bonferroni corrected.

Mixed ANOVA on time to fixate the target (when cued)

Source	***SS***	***df***	***MS***	***F***	***p***
Filtering type	1892483.119	2	946241.56	4.934	0.011
Error (Filtering)	9397727.141	49	191790.35		
Saliency	1369667.719	2	684833.86	13.17	0.000
Error (Saliency)	5096058.968	98	52000.602		
Filtering x Saliency	178743.848	4	44685.962	0.859	0.491

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	-222.1729	278.82227	-5.746	51	0.000
S1 v. Control	-167.1218	322.3777	-3.738	51	0.002
S5 v. Control	55.051	358.62694	1.107	51	0.860

NB. *p* value is Bonferroni corrected.

Post hoc independent samples *t* test between levels of filtering type

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Blur v. grey	-37.2593	96.97264	-0.384	33	1.000
Blur v. contrast	213.3686	78.59722	2.715	32	0.052
Grey v. contrast	250.6278	80.54564	3.112	33	0.015

NB. *p* value is Bonferroni corrected.

Mixed ANOVA on probability of fixating the region in target absent trials

Source	***SS***	***df***	***MS***	***F***	***p***
Filtering type	0.193	2	0.09662	3.629	0.034
Error (Filtering)	1.305	49	0.02662		
Saliency	0.65	2	0.325	84.328	0.000
Error (Saliency)	0.378	98	0.003853		
Filtering x Saliency	0.03329	4	0.008322	2.16	0.079

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	-0.0005	0.08134	-0.042	51	1.000
S1 v. Control	0.1367	0.10656	9.247	51	0.000
S5 v. Control	0.1371	0.07871	12.562	51	0.000

NB. *p* value is Bonferroni corrected.

Post hoc independent samples *t* test between levels of filtering type

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Blur v. grey	-0.0005	0.03475	-0.014	33	1.000
Blur v. contrast	0.0748	0.02952	2.534	32	0.075
Grey v. contrast	0.0753	0.03139	2.397	33	0.067

NB. *p* value is Bonferroni corrected.

Mixed ANOVA on hit rate

Source	***SS***	***df***	***MS***	***F***	***p***
Filtering type	0.165	2	0.08228	2.485	0.094
Error (Filtering)	1.622	49	0.03311		
Saliency	0.03363	2	0.01682	2.067	0.132
Error (Saliency)	0.797	98	0.008135		
Filtering x Saliency	0.04457	4	0.01114	1.37	0.25

Mixed ANOVA on d prime

Source	***SS***	***df***	***MS***	***F***	***p***
Filtering type	15.421	2	7.711	4.679	0.014
Error (Filtering)	80.741	49	1.648		
Saliency	2.013	2	1.006	1.936	0.150

Error (Saliency)	50.96	98	0.52		
Filtering x Saliency	2.936	4	0.734	1.411	0.236

Post hoc independent samples *t* test between levels of filtering type

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Blur v. grey	0.6771	0.26886	2.518	33	1.000
Blur v. contrast	0.0337	0.22665	0.149	32	0.028
Grey v. contrast	-0.6434	0.25661	-2.507	33	0.040

NB. *p* value is Bonferroni corrected.

Mixed ANOVA on RT

Source	***SS***	***df***	***MS***	***F***	***p***
Filtering type	35669308.88	2	17834654.44	5.944	0.005
Error (Filtering)	147019442.9	49	3000396.793		
Saliency	10730223.12	2	5365111.562	27.45	0.000
Error (Saliency)	19154427.4	98	195453.341		
Filtering x Saliency	1378348.421	4	344587.105	1.763	0.142

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	-575.3247	585.91365	-7.081	51	0.000
S1 v. Control	-540.5274	655.88398	-5.943	51	0.000
S5 v. Control	34.7972	659.0385	0.381	51	1.000

NB. *p* value is Bonferroni corrected.

Post hoc independent samples *t* test between levels of filtering type

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Blur v. grey	-228.7358	383.03025	-0.597	33	1.000
Blur v. contrast	-881.9289	303.71159	-2.904	32	0.040
Grey v. contrast	-1110.6647	325.64248	-3.411	33	0.006

NB. *p* value is Bonferroni corrected.

Mixed ANOVA on time to first fixate target (comparing with Exp. 4)

Source	***SS***	***df***	***MS***	***F***	***p***
Filtering type	3545674.542	3	1181891.514	7.535	0.000
Error (Filtering)	10039116.48	64	156861.195		
Saliency	1294675.579	2	647337.789	13.029	0.000
Error (Saliency)	6359476.337	128	49683.409		
Filtering x Saliency	297471.775	6	49578.629	0.998	0.430

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	-191.3636	284.64284	-5.544	67	0.000
S1 v. Control	-142.025	307.48404	-3.809	67	0.001
S5 v. Control	49.3386	350.00618	1.162	67	0.801

NB. *p* value is Bonferroni corrected.

Post hoc independent samples *t* test between levels of filtering type

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Blur v. grey	-37.2593	96.973	-0.384	33	1.000
Blur v. contrast	213.3686	78.59722	2.715	32	0.050
Blur v. normal	269.0813	75.96784	3.542	31	0.007
Grey v. contrast	250.6278	80.54564	3.112	33	0.011
Grey v. normal	306.3406	78.40015	3.907	32	0.001
Contrast v. normal	55.7127	49.78368	1.119	31	1.000

NB. *p* value is Bonferroni corrected.

Mixed ANOVA on RT (comparing with Exp. 4)

Source	***SS***	***df***	***MS***	***F***	***p***
Filtering type	45490415.76	3	15163471.92	5.541	0.002
Error (Filtering)	175152722	64	2736761.281		
Saliency	13833291.45	2	6916645.723	37.431	0.000
Error (Saliency)	23652099.52	128	184782.027		
Filtering x Saliency	1445401.599	6	240900.266	1.304	0.260

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	-559.8292	608.4866	-7.587	67	0.000
S1 v. Control	-549.3638	633.22262	-7.154	67	0.000
S5 v. Control	10.4655	593.75237	0.145	67	1.000

NB. *p* value is Bonferroni corrected.

Post hoc independent samples *t* test between levels of filtering type

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Blur v. grey	-228.7358	383.03	-0.597	33	1.000
Blur v. contrast	881.9289	303.712	2.904	32	0.054
Blur v. normal	726.4093	329.785	2.203	31	0.196
Grey v. contrast	1110.6647	325.64248	3.411	33	0.006
Grey v. normal	955.1451	350.47431	2.725	32	0.030
Contrast v. normal	-155.5196	250.67214	-0.62	31	1.000

NB. *p* value is Bonferroni corrected.

Mixed ANOVA on first fixation duration (comparing with Exp. 4)

Source	***SS***	***df***	***MS***	***F***	***p***
Filtering type	80974.036	3	26991.345	5.388	0.002
Error (Filtering)	320628.709	64	5009.824		
Saliency	19692.45	2	9846.225	13.166	0
Error (Saliency)	95723.003	128	747.836		
Filtering x Saliency	6994.259	6	1165.71	1.559	0.164

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	11.6376	33.85655	2.834	67	0.021
S1 v. Control	24.3136	39.29143	5.103	67	0.000
S5 v. Control	12.676	43.69437	2.392	67	0.058

NB. *p* value is Bonferroni corrected.

Post hoc independent samples *t* test between levels of filtering type

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Blur v. grey	-40.8256	14.10128	-2.895	33	0.026
Blur v. contrast	-26.0121	13.677	-1.902	32	0.409
Blur v. normal	8.8524	10.97553	0.807	31	1.000
Grey v. contrast	14.8135	16.24045	0.912	33	1.000
Grey v. normal	49.678	14.37308	3.456	32	0.005
Contrast v. normal	34.8645	13.91945	2.505	31	0.102

NB. *p* value is Bonferroni corrected.

Within-subject ANOVA on time to first fixate the target when cued
(split by change in saliency map)

Source	***SS***	***df***	***MS***	***F***	***p***
Correlation	594619.492	1	594619.492	2.062	0.17
Error (Correlation)	4614611.779	16	288413.236		
Saliency	1475715.245	1	1475715.245	7.649	0.014
Error (Saliency)	3086883.384	16	192930.212		
Correlation x Saliency	1074479.41	1	1074479.41	4.916	0.041
Error (Correlation x Saliency)	3497290.849	16	218580.678		

Experiment 5(b)

Within-subject ANOVA on RT

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	2540216.018	2	1270108.009	5.922	0.006
Error (Region type)	6863315.365	32	214478.605		
Inversion	534568.914	1	534568.914	1.637	0.219
Error (Inversion)	5223614	16	326475.875		
Region type x Inversion	102071.985	2	51035.992	0.256	0.776
Error (Region type x Inversion)	6380696.987	32	199396.781		

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	-173.2135	429.44726	-1.663	16	0.347
S1 v. Control	-385.8828	486.5616	-3.27	16	0.014
S5 v. Control	-212.6693	471.45379	-1.86	16	0.244

NB. *p* value is Bonferroni corrected.

Within-subject ANOVA on time to first fixate the target

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	206502.711	2	103251.355	0.687	0.51
Error (Region type)	4807734.584	32	150241.706		
Inversion	429136.307	1	429136.307	4.459	0.051
Error (Inversion)	1539786.301	16	96236.644		
Region type x Inversion	110317.544	2	55158.772	0.32	0.728
Error (Region type x Inversion)	5511766.595	32	172242.706		

Within-subject ANOVA on target absent fixation probability

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	0.395	2	0.198	67.4	0
Error (Region type)	0.094	32	0.003		
Inversion	0.037	1	0.037	13.765	0.002
Error (Inversion)	0.043	16	0.003		
Region type x Inversion	0.03	2	0.015	2.382	0.109
Error (Region type x Inversion)	0.199	32	0.006		

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	-0.0426	0.04988	-3.522	16	0.008
S1 v. Control	0.1055	0.05656	7.69	16	0.000
S5 v. Control	0.1481	0.05577	10.95	16	0.000

NB. *p* value is Bonferroni corrected.

Within-subject ANOVA on first fixation duration

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	5503.7	2	2751.85	3.663	0.037
Error (Region type)	24039.369	32	751.23		
Inversion	680.928	1	680.928	0.583	0.456
Error (Inversion)	18693.655	16	1168.353		
Region type x Inversion	2864.347	2	1432.174	1.552	0.227
Error (Region type x Inversion)	29527.964	32	922.749		

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	17.9714	22.92545	3.232	16	0.016
S1 v. Control	9.7482	28.46687	1.412	16	0.531
S5 v. Control	-8.2232	30.29442	-1.119	16	0.839

NB. *p* value is Bonferroni corrected.

Experiment 5(c)

Independent samples *t* test on RT between experiments

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Exp 4 v. 5(c)	-3121.4875	466.61971	-6.69	31	0.000
Exp 5(a) v. 5(c)	-2596.2687	442.02127	-5.874	33	0.000
Exp 5(b) v. 5(c)	-3271.2057	486.10009	-6.729	32	0.000

Within-subjects ANOVA on RT

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	8936854.909	2	4468427.454	3.614	0.038
Error (Region type)	39561941.08	32	1236310.659		

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	-1022.8461	1320.26221	-3.194	16	0.006
S1 v. Control	-573.7566	1961.27056	-1.206	16	0.245
S5 v. Control	449.0894	1352.10556	1.369	16	0.19

NB. *p* value is Bonferroni corrected.

Within-subjects ANOVA on first fixation time

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	2398272.919	2	1199136.459	3.389	0.046
Error (Region type)	11322905.09	32	353840.784		

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	-263.3584	910.1514	-1.193	16	0.25
S1 v. Control	267.8131	753.31809	1.466	16	0.162
S5 v. Control	531.1715	852.74907	2.568	16	0.021

NB. *p* value is Bonferroni corrected.

Within-subjects ANOVA on target absent fixation probability

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	0.163	2	0.081	21.609	0.000
Error (Region type)	0.12	32	0.004		

Post hoc paired samples *t* tests between levels of region type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
S1 v. S5	0.0243	0.08047	1.247	16	0.230
S1 v. Control	0.1301	0.085	6.31	16	0.000
S5 v. Control	0.1057	0.0942	4.628	16	0.000

NB. *p* value is Bonferroni corrected.

Within-subjects ANOVA on first fixation duration

Source	***SS***	***df***	***MS***	***F***	***p***
Region type	535.316	2	267.658	0.22	0.804
Error (Region type)	38908.963	32	1215.905		

Independent samples *t* test on fixation duration between experiments

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Exp 4 v. 5(c)	-29.2708	13.22589	-2.213	31	0.034

Experiment 6

Within subject ANOVA on accuracy

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	0.02722	2	0.01361	1.739	0.191
Error (Set size)	0.266	34	0.007827		
Saliency	0.002315	1	0.002315	0.339	0.568
Error (Saliency)	0.116	17	0.006825		
Set size x Saliency	0.0413	2	0.02065	3.115	0.057
Error (Set size x Saliency)	0.225	34	0.00629		

Within subject ANOVA on RT

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	2049038.403	2	1024519.201	14.736	0.000
Error (Set size)	2363823.434	34	69524.219		
Saliency	85182.367	1	85182.367	2.353	0.143
Error (Saliency)	615521.972	17	36207.175		
Set size x Saliency	338469.17	1.434	235976.303	4.041	0.042
Error (Set size x Saliency)	1423756.63	24.384	41875.195		

NB. Greenhouse-Geisser corrected

Post hoc paired samples *t* tests between levels of set size

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
8 v. 12	1.7333	290.10355	0.025	17	1.000
8 v. 16	-291.3222	212.16337	-5.826	17	0.000
12 v. 16	-293.0556	281.7788	-4.412	17	0.000

NB. *p* value is Bonferroni corrected.

Simple main effects

Source	***SS***	***df***	***MS***	***F***	***p***
Set size at non-salient	1965434	2	982717	14.13	0.0000
Set size at salient	422074	2	211037	3.035	0.0612
Error	2363823	34	69524		

Within subject ANOVA on time to target

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	1455709.36	2	727854.68	27.092	0.000
Error (Set size)	913451.455	34	26866.219		
Saliency	211.064	1	211.064	0.009	0.925
Error (Saliency)	391914.49	17	23053.794		
Set size x Saliency	132239.414	2	66119.707	1.39	0.263
Error (Set size x Saliency)	1617851.007	34	47583.853		

Post hoc paired samples *t* tests between levels of set size

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
8 v. 12	-149.0342	157.90646	-4.004	17	0.003
8 v. 16	-284.2697	199.5274	-6.045	17	0.000
12 v. 16	-135.2356	125.9088	-4.557	17	0.001

NB. *p* value is Bonferroni corrected.

Within subject ANOVA on first fixation duration

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	3322.968	2	1661.484	0.678	0.514
Error (Set size)	83344.329	34	2451.304		
Saliency	3316.793	1	3316.793	2.199	0.156
Error (Saliency)	25641.398	17	1508.318		
Set size x Saliency	8864.979	2	4432.49	1.824	0.177
Error (Set size x Saliency)	82632.441	34	2430.366		

Within subject ANOVA on first saccade amplitude

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	4.902	2	2.451	0.582	0.564
Error (Set size)	143.267	34	4.214		
Saliency	65.14	1	65.14	7.892	0.012
Error (Saliency)	140.313	17	8.254		
Set size x Saliency	3.045	2	1.522	0.618	0.545
Error (Set size x Saliency)	83.719	34	2.462		

Independent samples *t* test on distance from start point to target

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Salient v. Non-salient	1.394	1.744	0.639	58	0.427

One-sample *t* tests on string edit comparison with saliency

Condition	***M difference***	***t***	***df***	***p***
Non-salient 8	-0.0639	-6.7	17	0.000
Salient 8	0.0878	4.518	17	0.000
TA 8	-0.0328	-4.278	17	0.001
Non-salient 12	-0.0283	-2.47	17	0.024
Salient 12	0.0706	2.601	17	0.019
TA 12	0.0244	3.02	17	0.008
Non-salient 16	-0.0117	-1.338	17	0.198
Salient 16	0.0433	3.203	17	0.005
TA 16	-0.0039	-0.708	17	0.488

Test value=0

Within-subjects ANOVA on string edit comparison with saliency

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	0.0171	2	0.0086	2.8010	0.075
Error (Set size)	0.1040	34	0.0031		
Trial type	0.2950	2	0.1480	36.4700	0.000
Error (Trial type)	0.1380	34	0.0040		
Trial type x Set size	0.0560	4	0.0140	4.1270	0.005
Error (Trial type x Set size)	0.2310	68	0.0034		

Post hoc paired samples *t* tests between levels of trial type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Non v. salient	-0.1019	0.06547	-6.6	17	0.000
Non v. TA	-0.0306	0.03709	-3.495	17	0.008
TA v. salient	0.0713	0.04927	6.139	17	0.000

NB. *p* value is Bonferroni corrected.

Simple main effects

Source	***SS***	***df***	***MS***	***F***	***p***
Set size at non-salient	0.026	2	0.013	4.189	0.0237
Set size at salient	0.018	2	0.009	2.956	0.0655
Set size at target absent	0.029	2	0.015	4.819	0.0144
Error	0.104	34	0.003		

Post hoc paired samples *t* tests between levels of set size (at non salient)

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
8 v. 12	-0.0356	0.05501	-2.742	17	0.042
8 v. 16	-0.0522	0.05986	-3.701	17	0.005
12 v. 16	-0.0167	0.06362	-1.112	17	0.845

NB. *p* value is Bonferroni corrected.

Post hoc paired samples *t* tests between levels of set size (at target absent)

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
8 v. 12	-0.0572	0.04509	-5.384	17	0.000
8 v. 16	-0.0289	0.04324	-2.835	17	0.034
12 v. 16	0.0283	0.03312	3.629	17	0.006

NB. *p* value is Bonferroni corrected.

One-sample *t* tests on chance-adjusted within-subject similarity

Condition	***M difference***	***t***	***df***	***p***
Non-salient 8	-0.009	-1.271	17	0.221
Salient 8	-0.014	-1.761	17	0.096
TA 8	-0.0028	-0.664	17	0.516
Non-salient 12	-0.0072	-0.982	17	0.34
Salient 12	0.0072	1.093	17	0.29
TA 12	0.0111	2.938	17	0.009
Non-salient 16	0.0217	2.588	17	0.019
Salient 16	0.015	2.671	17	0.016
TA 16	0.0183	5.807	17	0.000

Test value=0

Within-subjects ANOVA on within-subject string edit comparison

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	0.0198	2	0.0099	12.9690	0.000
Error (Set size)	0.0259	34	0.0008		
Trial type	0.0016	2	0.0008	1.1190	0.338
Error (Trial type)	0.0248	34	0.0007		
Trial type x Set size	0.0033	4	0.0008	1.4460	0.228
Error (Trial type x Set size)	0.0382	68	0.0006		

Post hoc paired samples *t* tests between levels of set size

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
8 v. 12	-0.0124	0.02651	-1.985	17	0.19
8 v. 16	-0.027	0.02246	-5.107	17	0.000
12 v. 16	-0.0146	0.01783	-3.482	17	0.009

NB. *p* value is Bonferroni corrected.

One-sample *t* tests on chance-adjusted within-trial similarity

Condition	***M difference***	***t***	***df***	***p***
Non-salient 8	0.1133	11.682	17	0.000
Salient 8	0.1611	12.314	17	0.000
TA 8	0.0767	17.314	17	0.000
Non-salient 12	0.1072	16.226	17	0.000
Salient 12	0.1083	14.513	17	0.000
TA 12	0.0661	20.356	17	0.000
Non-salient 16	0.1006	15.031	17	0.000
Salient 16	0.0944	14.692	17	0.000
TA 16	0.0672	25.295	17	0.000

Test value=0

Within-subjects ANOVA on within-trial string edit comparison

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	0.0262	2	0.0131	11.7380	0.000
Error (Set size)	0.0380	34	0.0011		
Trial type	0.0757	2	0.0379	54.4460	0.000
Error (Trial type)	0.0236	34	0.0007		
Trial type x Set size	0.0210	4	0.0053	6.8400	0.000
Error (Trial type x Set size)	0.0382	68	0.0006		

Post hoc paired samples *t* tests between levels of set size

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
8 v. 12	0.0231	0.0306	3.209	17	0.015
8 v. 16	0.0296	0.03087	4.073	17	0.002
12 v. 16	0.0065	0.01852	1.485	17	0.468

NB. *p* value is Bonferroni corrected.

Post hoc paired samples *t* tests between levels of trial type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Non v. salient	-0.0143	0.02269	-2.666	17	0.049
Non v. TA	0.037	0.01576	9.974	17	0.000
TA v. salient	0.0513	0.02505	8.687	17	0.000

NB. *p* value is Bonferroni corrected.

Simple main effects

Source	***SS***	***df***	***MS***	***F***	***p***
Set size at non-salient	0.002	2	0.001	0.669	0.5187
Set size at salient	0.046	2	0.023	20.134	0
Set size at target absent	0.001	2	0.001	0.54	0.5878
Error	0.039	34	0.001		

Experiment 7

Within subject ANOVA on accuracy

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	0.0215	2	0.0107	2.7940	0.0770
Error (Set size)	0.1150	30	0.0038		
Saliency	0.0104	1	0.0104	2.9530	0.1060
Error (Saliency)	0.0529	15	0.0035		
Set size x Saliency	0.0002	2	0.0001	0.0230	0.9770
Error (Set size x Saliency)	0.1360	30	0.0045		

Within subject ANOVA on RT

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	229783.0690	2	114891.5350	5.6660	0.0080
Error (Set size)	608351.0700	30	20278.3690		
Saliency	5927.8340	1	5927.8340	0.5380	0.4750
Error (Saliency)	165401.0330	15	11026.7360		
Set size x Saliency	64618.8140	2	32309.4070	1.4130	0.2590
Error (Set size x Saliency)	686073.6580	30	22869.1220		

Post hoc paired samples *t* tests between levels of set size

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
8 v. 12	4.0515	134.97178	0.12	15	0.906
8 v. 16	-101.6987	153.62941	-2.648	15	0.018
12 v. 16	-105.7503	137.89754	-3.068	15	0.008

NB. *p* value is Bonferroni corrected.

Independent samples *t* test on RT between experiments

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Exp 6. v. 7	371.9127	92.06449	4.04	32	0.000

Independent samples *t* test on time to target between experiments

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Exp 6. v. 7	201.4552	45.150	4.462	32	0.000

Within subject ANOVA on time to target

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	296527.1820	2	148263.5910	19.5320	0.0000
Error (Set size)	227720.0270	30	7590.6680		
Saliency	3286.2960	1	3286.2960	0.3760	0.5490
Error (Saliency)	131149.5410	15	8743.3030		
Set size x Saliency	29366.4260	2	14683.2130	1.7080	0.1980
Error (Set size x Saliency)	257936.3610	30	8597.8790		

Post hoc paired samples *t* tests between levels of set size

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
8 v. 12	-73.9381	87.83041	-3.367	15	0.013
8 v. 16	-135.9619	73.68394	-7.381	15	0.000
12 v. 16	-62.0237	98.12491	-2.528	15	0.07

NB. *p* value is Bonferroni corrected.

Within subject ANOVA on first fixation duration

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	8819.3870	2	4409.6940	2.2190	0.1260
Error (Set size)	59619.3110	30	1987.3100		
Saliency	59.5730	1	59.5730	0.0280	0.8690
Error (Saliency)	31518.5080	15	2101.2340		
Set size x Saliency	1570.0130	2	785.0060	0.2420	0.7870
Error (Set size x Saliency)	97395.2260	30	3246.5080		

Within subject ANOVA on target saccadic amplitude

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	4.0540	2	2.0270	0.7430	0.4840
Error (Set size)	81.8460	30	2.7280		
Saliency	43.3620	1	43.3620	16.2440	0.0010
Error (Saliency)	40.0400	15	2.6690		
Set size x Saliency	4.7890	2	2.3940	1.1300	0.3360
Error (Set size x Saliency)	63.5420	30	2.1180		

Within-subjects ANOVA on string edit comparison with saliency

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	0.0120	2	0.0060	1.5800	0.2230
Error (Set size)	0.1190	30	0.0040		
Trial type	0.5660	2	0.2830	66.9680	0.0000
Error (Trial type)	0.1270	30	0.0040		
Set size x Trial type	0.2570	4	0.0640	13.2490	0.0000
Error (Set size x Trial type)	0.2910	60	0.0050		

Post hoc paired samples *t* tests between levels of trial type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Non-sal v. sal	-0.1502	0.06727	-8.932	15	0.000
Non-sal v. TA	-0.0473	0.03595	-5.262	15	0.000
Sal v. TA	0.1029	0.05136	8.015	15	0.000

NB. *p* value is Bonferroni corrected.

Simple main effects

Source	***SS***	***df***	***MS***	***F***	***p***
Set size at non-salient	0.041	2	0.021	5.243	0.0112
Set size at salient	0.222	2	0.111	28.051	0.000
Set size at target absent	0.006	2	0.003	0.802	0.4578
Error	0.119	30	0.004		

One-sample *t* tests on chance-adjusted observed v. saliency similarity

Condition	***M difference***	***t***	***df***	***p***
Non-salient 8	-0.0863	-8.594	15	0.000
Salient 8	0.2	7.609	15	0.000
TA 8	-0.015	-1.098	15	0.290
Non-salient 12	-0.0169	-1.526	15	0.148
Salient 12	0.05	1.966	15	0.068
TA 12	0.0094	0.896	15	0.385
Non-salient 16	-0.035	-4.583	15	0.000
Salient 16	0.0625	2.484	15	0.025
TA 16	0.0094	1.802	15	0.092

Test value=0

Within-subjects ANOVA on within-subject string edit comparison

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	0.0370	2	0.0180	15.0320	0.0000
Error (Set size)	0.0370	30	0.0010		
Trial type	0.0000	2	0.0000	0.0060	0.9940
Error (Trial type)	0.0310	30	0.0010		
Set size x Trial type	0.0150	4	0.0040	3.4560	0.0130
Error (Set size x Trial type)	0.0640	60	0.0010		

Post hoc paired samples *t* tests between levels of set size

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
8 v. 12	-0.0288	0.0219	-5.25	15	0.000
8 v. 16	-0.0375	0.0308	-4.869	15	0.001
12 v. 16	-0.0087	0.03208	-1.091	15	0.878

NB. *p* value is Bonferroni corrected.

Simple main effects

Source	***SS***	***df***	***MS***	***F***	***p***
Set size at non-salient	0.014	2	0.007	5.819	0.0073
Set size at salient	0.032	2	0.016	12.862	0.0001
Set size at target absent	0.006	2	0.003	2.351	0.1126
Error	0.037	30	0.001		

One-sample *t* tests on within-subject similarity

Condition	***M difference***	***t***	***df***	***p***
Non-salient 8	-0.0119	-0.879	15	0.393
Salient 8	-0.0306	-3.629	15	0.002
TA 8	-0.0106	-2.262	15	0.039
Non-salient 12	-0.0038	-0.415	15	0.684
Salient 12	0.0294	3.179	15	0.006
TA 12	0.0075	2.423	15	0.029
Non-salient 16	0.0281	2.679	15	0.017
Salient 16	0.0156	2.157	15	0.048
TA 16	0.0156	4.948	15	0.000

Test value=0

Within-subjects ANOVA on within-trial string edit comparison

Source	***SS***	***df***	***MS***	***F***	***p***
Set size	0.0790	2	0.0400	29.0110	0.0000
Error (Set size)	0.0410	30	0.0010		
Trial type	0.3710	2	0.1850	212.0040	0.0000
Error (Trial type)	0.0260	30	0.0010		
Set size x Trial type	0.0320	4	0.0080	6.2760	0.0000
Error (Set size x Trial type)	0.0760	60	0.0010		

Post hoc paired samples *t* tests between levels of set size

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
8 v. 12	0.0477	0.03042	6.273	15	0.000
8 v. 16	0.0517	0.02865	7.214	15	0.000
12 v. 16	0.004	0.03146	0.503	15	1.000

NB. *p* value is Bonferroni corrected.

Post hoc paired samples *t* tests between levels of trial type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Non-sal v. sal	-0.02	0.0281	-2.847	15	0.037
Non-sal v. TA	0.0963	0.02076	18.546	15	0.000
Sal v. TA	0.1163	0.02299	20.223	15	0.000

NB. *p* value is Bonferroni corrected.

Simple main effects

Source	***SS***	***df***	***MS***	***F***	***p***
Set size at non-salient	0.032	2	0.016	11.815	0.0002
Set size at salient	0.076	2	0.038	27.896	0.0000
Set size at target absent	0.003	2	0.001	0.933	0.4043
Error	0.041	30	0.001		

One-sample *t* tests on within-trial similarity

Condition	***M difference***	***t***	***df***	***p***
Non-salient 8	0.215	14.606	15	0.000
Salient 8	0.255	25.5	15	0.000
TA 8	0.0906	18.942	15	0.000
Non-salient 12	0.1544	16.2	15	0.000
Salient 12	0.1781	16.537	15	0.000
TA 12	0.085	16.724	15	0.000
Non-salient 16	0.1681	14.977	15	0.000
Salient 16	0.1644	14.904	15	0.000
TA 16	0.0731	16.448	15	0.000

Test value=0

Experiment 8

Within subject ANOVA on response time

Source	***SS***	***df***	***MS***	***F***	***p***
Position	433972.2460	1	433972.2460	10.1340	0.0050
Error (Position)	813638.8380	19	42823.0970		
Angle	148569.1430	2	74284.5720	2.2670	0.1170
Error (Angle)	1245251.5030	38	32769.7760		
Position x Angle	297568.3310	2	148784.1660	3.0410	0.0600
Error (Position x Angle)	1858911.4430	38	48918.7220		

Within subject ANOVA on time to target

Source	***SS***	***df***	***MS***	***F***	***p***
Position	217464.5880	1	217464.5880	3.8990	0.0630
Error (Position)	1059601.8040	19	55768.5160		
Angle	240540.3580	2	120270.1790	1.8170	0.1760
Error (Angle)	2515132.0280	38	66187.6850		
Position x Angle	514729.6870	2	257364.8440	4.4410	0.0180
Error (Position x Angle)	2201947.1890	38	57945.9790		

Simple main effects

Source	***SS***	***df***	***MS***	***F***	***p***
Position at 0	190127	1	190127	3.409	0.0805
Position at 90	451782	1	451782	8.101	0.0103
Position at 180	90284	1	90284	1.619	0.2186
Error	1059602	19	55769		

Within subject ANOVA on distractor fixation probability

Source	***SS***	***df***	***MS***	***F***	***p***
Position	0.021	1	0.021	0.488	0.4932
Error (Position)	0.813	19	0.043		
Angle	2.879	2	1.439	44.126	0.0000
Error (Angle)	1.240	38	0.033		
Position x Angle	0.206	2	0.103	1.498	0.2365
Error (Position x Angle)	2.607	38	0.069		

Post hoc paired samples *t* tests between levels of angle

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
0 v. 90	0.3312	0.21197	6.989	19	0.000
0 v. 180	0.3258	0.18289	7.967	19	0.000
90 v. 180	-0.0054	0.13958	-0.174	19	1.000

NB. *p* value is Bonferroni corrected.

Within subject ANOVA on response time

Source	***SS***	***df***	***MS***	***F***	***p***
Distractor fixation	1759732.5760	1	1759732.5760	9.4730	0.0060
Error (D. Fix.)	3529411.5030	19	185758.5000		
Position	662959.0710	1	662959.0710	6.0090	0.0240
Error (Position)	2096333.5370	19	110333.3440		
Angle	276762.2830	2	138381.1410	1.9180	0.1610
Error (Angle)	2742216.3440	38	72163.5880		
D. Fix x Position	1377087.0120	1	1377087.0120	12.9690	0.0020
Error	2017399.7710	19	106178.9350		
D. Fix x Angle	1764430.5540	2	882215.2770	5.6870	0.0070
Error	5894707.0030	38	155123.8680		
Position x angle	1339535.9340	2	669767.9670	9.5230	0.0000
Error	2672710.6810	38	70334.4920		
D. Fix x position x angle	798711.8610	2	399355.9310	5.5780	0.0080
Error	2720663.9650	38	71596.4200		

Simple main effects

Source	***SS***	***df***	***MS***	***F***	***p***
Position at fixated	1975494	1	1975494	17.91	0.0005
Position at not fixated	64538	1	64538	0.585	0.4538
Error	2096341	19	110334		
Angle at fixated	513429	2	256714	3.557	0.0384
Angle at not fixated	1527764	2	763882	10.585	0.0002
Error	2742211	38	72163		
Position at 0	70058	1	70058	0.635	0.4354
Position at 90	104226	1	104226	0.945	0.3433
Position at 180	1828208	1	1828208	16.570	0.0007
Error	2096342	19	110333		
D. Fix. at edge 0	15505.088	1	15505.088	0.083	0.7758
D. Fix. at edge 90	60164.068	1	60164.068	0.324	0.5759
D. Fix. at edge 180	33250.216	1	33250.216	0.179	0.677
D. Fix. at centre 0	2354471.824	1	2354471.824	12.675	0.0021
D. Fix. at centre 90	3172252.168	1	3172252.168	17.077	0.0006
D. Fix. at centre 180	64313.984	1	64313.984	0.346	0.5632
Error	3529396.82	19	185757.727		

Experiment 9

Within subject ANOVA on hit rate

Source	***SS***	***df***	***MS***	***F***	***p***
Eccentricity	0.0056	1	0.0056	1.0470	0.3230
Error (Ecc.)	0.0806	15	0.0054		
Saliency	0.0025	1	0.0025	0.4480	0.5140
Error (Saliency)	0.0838	15	0.0056		
Ecc. x Saliency	0.0014	1	0.0014	0.6050	0.4490
Error (Ecc. x Saliency)	0.0348	15	0.0023		

Within subject ANOVA on RT

Source	***SS***	***df***	***MS***	***F***	***p***
Eccentricity	98051.7720	1	98051.7720	27.0850	0.0000
Error (Ecc.)	54302.7050	15	3620.1800		
Saliency	25387.2950	1	25387.2950	8.6300	0.0100
Error (Saliency)	44127.2670	15	2941.8180		
Ecc. x Saliency	6023.0520	1	6023.0520	1.7580	0.2050
Error (Ecc. x Saliency)	51392.2160	15	3426.1480		

Paired *t* test on RT

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Target absent v. target present	168.3044	70.72689	9.519	15	0.0000

Within subject ANOVA on time to target

Source	***SS***	***df***	***MS***	***F***	***p***
Eccentricity	359060.0380	1	359060.0380	171.8700	0.0000
Error (Ecc.)	29247.9240	14	2089.1370		
Saliency	15083.2530	1	15083.2530	16.5450	0.0010
Error (Saliency)	12762.9220	14	911.6370		
Ecc. x Saliency	3150.7620	1	3150.7620	2.2810	0.1530
Error (Ecc. x Saliency)	19341.8100	14	1381.5580		

Within subject ANOVA on probability of distractor fixation

Source	***SS***	***df***	***MS***	***F***	***p***
Eccentricity	2.0720	1	2.0720	108.5110	0.0000
Error (Ecc.)	0.2670	14	0.0191		
Saliency	0.0000	1	0.0000	0.0130	0.9100
Error (Saliency)	0.0443	14	0.0032		
Ecc. x Saliency	0.0220	1	0.0220	4.1240	0.0620
Error (Ecc. x Saliency)	0.0748	14	0.0053		

Within subject ANOVA on initial saccade latency

Source	***SS***	***df***	***MS***	***F***	***p***
Eccentricity	71.1340	1	71.1340	0.1980	0.6630
Error (Ecc.)	5029.4620	14	359.2470		
Saliency	151.6490	1	151.6490	0.3450	0.5660
Error (Saliency)	6156.4790	14	439.7490		
Ecc. x Saliency	58.6940	1	58.6940	0.2010	0.6610
Error (Ecc. x Saliency)	4095.9140	14	292.5650		

Within subject ANOVA on probability of inspection

Source	***SS***	***df***	***MS***	***F***	***p***
Eccentricity	2.7400	1	2.7400	149.1790	0.0000
Error (Ecc.)	0.2570	14	0.0184		
Saliency rank	0.1060	4	0.0265	6.4710	0.0000
Error (Saliency)	0.2290	56	0.0041		
Ecc. x Saliency	0.0594	4	0.0149	2.8760	0.0310
Error (Ecc. x Saliency)	0.2890	56	0.0052		

Within subject ANOVA on mean horizontal landing position

Source	***SS***	***df***	***MS***	***F***	***p***
Eccentricity	0.0148	1	0.0148	0.5180	0.4840
Error (Ecc.)	0.4000	14	0.0285		
Saliency	0.0610	1	0.0610	1.0740	0.3180
Error (Saliency)	0.7960	14	0.0568		
Ecc. x Saliency	0.0151	1	0.0151	0.4070	0.5340
Error (Ecc. x Saliency)	0.5200	14	0.0371		

Within subject ANOVA on mean vertical landing position

Source	***SS***	***df***	***MS***	***F***	***p***
Eccentricity	0.3670	1	0.3670	10.2600	0.0060
Error (Ecc.)	0.5010	14	0.0358		
Saliency	0.0561	1	0.0561	0.8890	0.3620
Error (Saliency)	0.8840	14	0.0631		
Ecc. x Saliency	0.0038	1	0.0038	0.1910	0.6690
Error (Ecc. x Saliency)	0.2810	14	0.0201		

Within subject ANOVA on refixation probability

Source	***SS***	***df***	***MS***	***F***	***p***
Landing position	0.3710	3	0.1240	5.6300	0.0020
Error (Landing position.)	0.9220	42	0.0220		

Within subject ANOVA on fixation duration

Source	***SS***	***df***	***MS***	***F***	***p***
Landing position	4518.4880	3	1506.1630	0.6560	0.5840
Error (Landing position.)	96488.4780	42	2297.3450		

Experiment 10(a)

Paired *t* test on frequency of left and right saccades

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Left v.Right	0.3077	17.72656	0.063	12	0.951

Within subject ANOVA on saccade frequency

Source	***SS***	***df***	***MS***	***F***	***p***
Picture orientation	14.9080	4	3.7270	2.1070	0.0940
Error (Pic. Or.)	84.8920	48	1.7690		
Axis	431.1690	3	143.7230	3.9700	0.0150
Error (axis)	1303.2310	36	36.2010		
Pic. Or. x Axis	5262.2920	12	438.5240	21.2450	0.0000
Error (Pic. Or. x Axis)	2972.3080	144	20.6410		

Post hoc paired *t* tests between levels of axis

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
0˚ v.45˚	1.7077	3.01951	2.039	12	0.385
0˚ v.90˚	2.5385	5.49508	1.666	12	0.730
0˚ v.135˚	3.5077	3.53328	3.579	12	0.023
45˚ v.90˚	0.8308	4.22125	0.71	12	1.000
45˚ v.135˚	1.8	2.23159	2.908	12	0.079
90˚ v.135˚	0.9692	3.50509	0.997	12	1.000

NB. *p* value is Bonferroni corrected.

Planned *t* tests between axis of orientation and elsewhere

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
At 0˚ (0˚ v.*M*)	12.1282	6.61443	6.611	12	0.000
At 45˚ (45˚ v.*M*)	12.4872	8.26597	5.447	12	0.000
At 90˚ (90˚ v.*M*)	10.6923	6.74051	5.719	12	0.000
At 135˚ (135˚ v.*M*)	7.1538	7.79793	3.308	12	0.006
At 180˚ (180˚ v.*M*)	8.2051	5.46629	5.412	12	0.000

Within subject ANOVA on saccade frequency over time

Source	***SS***	***df***	***MS***	***F***	***p***
Saccade number	0.3620	4	0.0904	2.4960	0.0550
Error (Sac. Num.)	1.7380	48	0.0362		
Axis	1356.2620	3	452.0870	33.3750	0.0000
Error (axis)	487.6380	36	13.5460		
Sac Num x Axis	282.6230	12	23.5520	2.3740	0.0080
Error (Sac Num. x Axis)	1428.4770	144	9.9200		

Post hoc paired *t* tests between levels of axis

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
0˚ v.45˚	5.0154	2.70242	6.691	12	0.000
0˚ v.90˚	5.3077	3.27884	5.837	12	0.000
0˚ v.135˚	5.4615	2.47303	7.963	12	0.000
45˚ v.90˚	0.2923	1.98262	0.532	12	1.000
45˚ v.135˚	0.4462	1.66964	0.963	12	1.000
90˚ v.135˚	0.1538	1.2732	0.436	12	1.000

NB. *p* value is Bonferroni corrected.

Planned *t* tests between 0˚ and elsewhere

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
At 2nd (0˚ v.*M*)	1.7436	4.26474	1.474	12	0.166
At 3rd (45˚ v.*M*)	6.7692	4.75976	5.128	12	0.000
At 4th (90˚ v.*M*)	4.9231	5.05919	3.509	12	0.004
At 5th (135˚ v.*M*)	5.7692	4.59747	4.524	12	0.001
At 6th (180˚ v.*M*)	7.1026	4.42716	5.784	12	0.000

Within subject ANOVA on saccade amplitude

Source	***SS***	***df***	***MS***	***F***	***p***
Picture orientation	15.2000	4	3.8000	2.4690	0.0570
Error (Pic. Or.)	73.8680	48	1.5390		
Axis	15.5470	3	5.1820	2.2890	0.0950
Error (axis)	81.5160	36	2.2640		
Pic. Or. x Axis	186.7230	12	15.5600	8.5890	0.0000
Error (Pic. Or. x Axis)	260.8770	144	1.8120		

Planned *t* tests between axis of orientation and elsewhere

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
At 0˚ (0˚ v.*M*)	2.0362	1.2983	5.655	12	0.0000
At 45˚ (45˚ v.*M*)	2.9392	1.10912	9.555	12	0.0000
At 90˚ (90˚ v.*M*)	1.0708	1.1724	3.293	12	0.0060
At 135˚ (135˚ v.*M*)	1.2262	1.85773	2.38	12	0.0350
At 180˚ (180˚ v.*M*)	0.7577	1.14994	2.376	12	0.0350

Experiment 10(b)

Paired *t* tests on saliency in different images

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Landscapes (middle v. outer)	2.3175	3.1143	4.402	39	0.000
Interiors (middle v. outer)	0.7176	2.91301	1.349	39	0.188

Paired *t* tests on edge content in different images

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Landscapes (horizontal v. vertical)	52119.475	66014.36881	4.993	39	0.000
Interiors (horizontal v. vertical)	8239.35	31311.82445	1.664	39	0.104

Within subject ANOVA on hit rate

Source	***SS***	***df***	***MS***	***F***	***p***
Picture type	0.0014	1	0.0014	0.0210	0.8870
Error (Pic. Type)	0.7490	11	0.0681		
Picture orientation	2.6440	4	0.6610	17.7620	0.0000
Error (Pic. Or.)	1.6380	44	0.0372		
Pic. Type x Pic Or	0.3180	4	0.0794	1.8260	0.1410
Error (Pic. Type x Pic Or)	1.9140	44	0.0435		

Post hoc paired *t* tests between levels of picture orientation

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
0˚ v.45˚	0.4132	0.18588	7.7	11	0.000
0˚ v.90˚	0.2813	0.16101	6.051	11	0.001
0˚ v.135˚	0.1389	0.12479	3.855	11	0.027
0˚ v.180˚	0.3438	0.20345	5.853	11	0.001
45˚ v.90˚	-0.1319	0.19851	-2.303	11	0.418
45˚ v.135˚	-0.2743	0.20448	-4.647	11	0.007
45˚ v.180˚	-0.0694	0.26905	-0.894	11	1.000
90˚ v.135˚	-0.1424	0.1652	-2.985	11	0.124
90˚ v.180˚	0.0625	0.21504	1.007	11	1.000
135˚ v.180˚	0.2049	0.16615	4.271	11	0.013

NB. *p* value is Bonferroni corrected.

One-sample *t* tests on hit rate

Orientation	***M difference***	***t***	***df***	***p***
0˚	0.3958	15.284	11	0.000
45˚	-0.0174	-0.277	11	0.787
90˚	0.1146	2.561	11	0.026
135˚	0.2569	5.144	11	0.000
180˚	0.0521	0.877	11	0.399

Test value=0.5

Within subject ANOVA on orientation recognition rate

Source	***SS***	***df***	***MS***	***F***	***p***
Picture type	0.0306	1	0.0306	0.9600	0.3480
Error (Pic. Type)	0.3510	11	0.0319		
Picture orientation	6.4310	4	1.6080	28.0900	0.0000
Error (Pic. Or.)	2.5190	44	0.0572		
Pic. Type x Pic Or	0.3930	4	0.0983	3.4420	0.0160
Error (Pic. Type x Pic Or)	1.2570	44	0.0286		

Post hoc paired *t* tests between levels of picture orientation

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
0˚ v.45˚	0.6215	0.3215	6.697	11	0.000
0˚ v.90˚	0.5938	0.31589	6.511	11	0.000
0˚ v.135˚	0.5729	0.28434	6.98	11	0.000
0˚ v.180˚	0.4271	0.29469	5.02	11	0.004
45˚ v.90˚	-0.0278	0.19651	-0.49	11	1.000
45˚ v.135˚	-0.0486	0.11491	-1.465	11	1.000
45˚ v.180˚	-0.1944	0.25644	-2.627	11	0.235
90˚ v.135˚	-0.0208	0.15841	-0.456	11	1.000
90˚ v.180˚	-0.1667	0.14543	-3.97	11	0.022
135˚ v.180˚	-0.1458	0.19422	-2.601	11	0.246

NB. *p* value is Bonferroni corrected.

Simple main effects

Source	***SS***	***df***	***MS***	***F***	***p***
Pic type at 0˚	0.003	1	0.003	0.082	0.7802
Pic type at 45˚	0.005	1	0.005	0.143	0.713
Pic type at 90˚	0.000	1	0.000	0.000	1.000
Pic type at 135˚	0.042	1	0.042	1.309	0.2769
Pic type at 180˚	0.375	1	0.375	11.779	0.0056
Error	0.35	11	0.032		

One-sample *t* tests on orientation recognition rate

Orientation	***M difference***	***t***	***df***	***p***
0˚	0.4771	6.27	11	0.000
45˚	-0.1444	-4.329	11	0.001
90˚	-0.1167	-3.283	11	0.007
135˚	-0.0958	-3.7	11	0.004
180˚	0.05	1.039	11	0.321

Test value=0.2

Within subject ANOVA on saccade frequency (in landscapes)

Source	***SS***	***df***	***MS***	***F***	***p***
Pic orientation	6.0250	4	1.5060	1.2950	0.2870
Error (Pic. Ori.)	51.1750	44	1.1630		
Axis	454.3120	3	151.4370	8.6920	0.0000
Error (axis)	574.9380	33	17.4220		
Pic Ori x Axis	1406.2080	12	117.1840	13.6190	0.0000
Error (Pic Ori. x Axis)	1135.7920	132	8.6040		

Post hoc paired *t* tests between levels of axis

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
0˚ v.45˚	-0.5667	1.67947	-1.169	11	1.000
0˚ v.90˚	-3.2167	3.46878	-3.212	11	0.050
0˚ v.135˚	0.2333	0.95664	0.845	11	1.000
45˚ v.90˚	-2.65	3.18648	-2.881	11	0.090
45˚ v.135˚	0.8	1.94282	1.426	11	1.000
90˚ v.135˚	3.45	3.48099	3.433	11	0.034

NB. *p* value is Bonferroni corrected.

Planned *t* tests between 0˚ and elsewhere

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
At 0˚ (0˚ v.*M*)	2.0833	3.2353	2.231	11	0.047
At 45˚ (45˚ v.*M*)	7.1667	3.27679	7.576	11	0.000
At 90˚ (90˚ v.*M*)	8.6389	5.31333	5.632	11	0.000
At 135˚ (135˚ v.*M*)	4.1389	3.83355	3.74	11	0.003
At 180˚ (180˚ v.*M*)	0.6111	3.20931	0.66	11	0.523

Within subject ANOVA on saccade frequency (in interiors)

Source	***SS***	***df***	***MS***	***F***	***p***
Pic orientation	10.3500	4	2.5870	2.2570	0.0780
Error (Pic. Ori.)	50.4500	44	1.1470		
Axis	40.7670	3	13.5890	0.9280	0.4380
Error (axis)	483.2330	33	14.6430		

Pic Ori x Axis	907.8170	12	75.6510	10.9470	0.0000
Error (Pic Ori. x Axis)	912.1830	132	6.9100		

Planned *t* tests between 0˚ and elsewhere

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
At 0˚ (0˚ v.*M*)	3.9167	2.17016	6.252	11	0.000
At 45˚ (45˚ v.*M*)	1.25	3.63242	1.192	11	0.258
At 90˚ (90˚ v.*M*)	4.25	3.55086	4.146	11	0.002
At 135˚ (135˚ v.*M*)	3.5833	4.75697	2.609	11	0.024
At 180˚ (180˚ v.*M*)	2.3333	3.67904	2.197	11	0.050

Within subject ANOVA on saccade frequency at encoding by picture type

Source	***SS***	***df***	***MS***	***F***	***p***
Pic type	2.0420	1	2.0420	0.1170	0.7390
Error (Pic. Type.)	191.7080	11	17.4280		
Axis	7570.4170	3	2523.4720	40.3400	0.0000
Error (axis)	2064.3330	33	62.5560		
Pic Type x Axis	331.0420	3	110.3470	2.8600	0.0520
Error (Pic Type. x Axis)	1273.2080	33	38.5820		

Post hoc paired *t* tests between levels of axis

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
0˚ v.45˚	20.75	4.42873	16.23	11	0.000

0˚ v.90˚	10.4583	10.88255	3.329	11	0.040
0˚ v.135˚	21.9583	4.47446	17	11	0.000
45˚ v.90˚	-10.2917	9.97373	-3.575	11	0.026
45˚ v.135˚	1.2083	5.49983	0.761	11	1.000
90˚ v.135˚	11.5	9.35657	4.258	11	0.008

NB. *p* value is Bonferroni corrected.

Simple main effects

Source	***SS***	***df***	***MS***	***F***	***p***
Pic type at 0˚	210.042	1	210.042	12.052	0.0052
Pic type at 45˚	45.375	1	45.375	2.604	0.1349
Pic type at 90˚	73.5	1	73.5	4.217	0.0646
Pic type at 135˚	4.167	1	4.167	0.239	0.6345
Error	191.708	11	17.428		

Within subject ANOVA on saccade frequency at test

Source	***SS***	***df***	***MS***	***F***	***p***
Exposure	6.5100	1	6.5100	0.1960	0.6670
Error (Exposure)	365.3650	11	33.2150		
Axis	17596.1980	3	5865.3990	5.0370	0.0060
Error (axis)	38429.9270	33	1164.5430		
Exposure x Axis	3045.2810	3	1015.0940	19.5170	0.0000
Error (Exposure x Axis)	1716.3440	33	52.0100		

Post hoc paired *t* tests between levels of axis

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
0˚ v.45˚	28.375	20.50624	4.793	11	0.003
0˚ v.90˚	7.4583	50.34809	0.513	11	1.000
0˚ v.135˚	31.9583	30.5625	3.622	11	0.024
45˚ v.90˚	-20.9167	36.80775	-1.969	11	0.448
45˚ v.135˚	3.5833	20.71872	0.599	11	1.000
90˚ v.135˚	24.5	36.24475	2.342	11	0.234

NB. *p* value is Bonferroni corrected.

Simple main effects

Source	***SS***	***df***	***MS***	***F***	***p***
Exposure at 0˚	32.667	1	32.667	0.983	0.3427
Exposure at 45˚	1276.042	1	1276.042	38.418	0.0001
Exposure at 90˚	1717.042	1	1717.042	51.695	0.0000
Exposure at 135˚	26.042	1	26.042	0.784	0.3949
Error	365.365	11	33.215		

Paired *t* test on proportion of vertical saccades

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Hits v.FAs	0.0613	0.05955	3.565	11	0.004

Experiment 11(a)

Mixed ANOVA on RT

Source	***SS***	***df***	***MS***	***F***	***p***
Group	6239502.5450	1	6239502.5450	27.7090	0.0000
Error (Group)	5179068.4320	23	225176.8880		
Saliency	534877.3910	1	534877.3910	19.7720	0.0000
Group x Saliency	182820.1720	1	182820.1720	6.7580	0.0160
Error (Saliency)	622188.4360	23	27051.6710		

Post hoc paired *t* tests between levels of saliency

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Low v. Med (YCs)	87.6933	219.93329	1.544	14	0.145
Low v. Med (AMCs)	334.5537	251.03956	4.214	9	0.002

Mixed ANOVA on number of fixations

Source	***SS***	***df***	***MS***	***F***	***p***
Group	452.6460	1	452.6460	16.6150	0.0000
Error (Group)	626.5940	23	27.2430		
Trial type	566.6360	2	283.3180	36.8850	0.0000
Group x Trial type	77.8660	2	38.9330	5.0690	0.0100
Error (Trial type)	353.3300	46	7.6810		

Post hoc paired *t* tests between levels of trial type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Low v. Med	0.8695	0.77225	5.629	24	0.000
Low v. TA	-5.0235	5.02769	-4.996	24	0.000
Med v. TA	-5.8929	5.29391	-5.566	24	0.000

NB. *p* value is Bonferroni corrected.

Independent samples *t* test on fixation duration

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
AMCs v. YCs	21.1695	11.35406	1.864	23	0.075

Independent samples *t* test on saccade amplitude

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
AMCs v. YCs	0.126	0.59194	0.213	23	0.833

Mixed ANOVA on proportion of trials where target was fixated

Source	***SS***	***df***	***MS***	***F***	***p***
Group	0.1840	1	0.1840	2.3280	0.1410
Error (Group)	1.8190	23	0.0791		
Saliency	0.0353	1	0.0353	4.5540	0.0440
Group x Saliency	0.0002	1	0.0002	0.0280	0.8690
Error (Saliency)	0.1780	23	0.0078		

Mixed ANOVA on ordinal fixation on target

Source	***SS***	***df***	***MS***	***F***	***p***
Group	10.2650	1	10.2650	3.1670	0.0880
Error (Group)	74.5520	23	3.2410		
Saliency	9.4410	1	9.4410	10.0660	0.0040
Group x Saliency	0.5320	1	0.5320	0.5670	0.4590
Error (Saliency)	21.5720	23	0.9380		

Independent samples *t* test on proportion of trials where most salient region was fixated

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
AMCs v. YCs	0.1377	0.05192	2.651	23	0.014

Mixed ANOVA on proportion of fixations on target

Source	***SS***	***df***	***MS***	***F***	***p***
Group	0.0029	1	0.0029	0.1220	0.7300
Error (Group)	0.5440	23	0.0237		
Saliency	0.0435	1	0.0435	10.5410	0.0040
Group x Saliency	0.0020	1	0.0020	0.4900	0.4910
Error (Saliency)	0.0948	23	0.0041		

Independent samples *t* test on proportion of trials where most salient region was fixated

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
AMCs v. YCs	0.0056	0.01053	0.528	23	0.603

Independent samples *t* test on saliency at fixation

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
AMCs v. YCs	0.07	0.07126	0.982	23	0.336

Experiment 11(b)

Between-groups ANOVA on d prime

Source	***SS***	***df***	***MS***	***F***	***p***
Scene type	9.236	1	9.236	28.634	.000
Error (Scene type)	12.579	39	.323		

Mixed ANOVA on number of fixations

Source	***SS***	***df***	***MS***	***F***	***p***
Scene type	17.431	1	17.431	2.128	0.153
Error (Scene type)	319.481	39	8.192		
Trial type	25.747	2	12.873	12.988	0.000
Scene type x Trial type	1.397	2	0.699	0.705	0.497
Error (Trial type)	77.311	78	0.991		

Post hoc paired *t* tests between levels of trial type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Enc v. Old	-1.029	1.64764	-3.999	40	0.001
Enc v. New	-0.8586	1.7444	-3.152	40	0.006
Old v. New	0.1704	0.38142	2.861	40	0.077

NB. *p* value is Bonferroni corrected.

Mixed ANOVA on fixation duration

Source	***SS***	***df***	***MS***	***F***	***p***
Scene type	12983.352	1	12983.352	1.742	0.195
Error (Scene type)	290604.513	39	7451.398		
Trial type	14037.243	2	7018.621	8.61	0.000
Scene type x Trial type	4400.463	2	2200.231	2.699	0.074
Error (Trial type)	63586.103	78	815.206		

Post hoc paired *t* tests between levels of trial type

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
Enc v. Old	23.6707	45.83415	3.307	40	0.002
Enc v. New	13.4722	51.79011	1.666	40	0.104
Old v. New	-10.1985	17.7766	-3.674	40	0.001

NB. *p* value is Bonferroni corrected.

Mixed ANOVA on saccade amplitude

Source	***SS***	***df***	***MS***	***F***	***p***
Scene type	67.813	1	67.813	14.31	0.001
Error (Scene type)	184.819	39	4.739		
Trial type	6.992	2	3.496	15.833	0.000
Scene type x Trial type	4.213	2	2.106	9.54	0.000
Error (Trial type)	17.222	78	0.221		

Simple main effects

Source	***SS***	***df***	***MS***	***F***	***p***
Scene type at Enc	10.63	1	10.63	6.157	0.0145
Scene type at Old	37.938	1	37.938	21.970	0.000
Scene type at new	23.455	1	23.455	13.582	0.000
Error	202.042	117	1.727		

Independent samples *t* test on proportion of fixations in salient regions

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Scenes v. Fractals	0.0138	0.00712	1.939	39	0.06

Independent samples *t* test on saliency at fixation

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Scenes v. Fractals	0.1711	0.02736	6.252	39	0.000

Independent samples *t* test on distance from fixations to centre

	Differences				
Comparison	***M***	***SEM***	***t***	***df***	***p***
Scenes v. Fractals	1.2488	0.240	5.298	39	0.000

Within subject ANOVA on saccade frequency in fractals

Source	***SS***	***df***	***MS***	***F***	***p***
Axis	194744.977	3	64914.992	6.092	0.002
Error (axis)	319670.773	30	10655.692		

Planned *t* test between 0˚ and elsewhere

	Paired differences				
Comparison	***M***	***SD***	***t***	***df***	***p***
0˚ v.*M*	147	157.79318	3.09	10	0.011